The first complete chloroplast genome of the Genistoid legume Lupinus luteus: evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family

Guillaume E. Martin, Mathieu Rousseau-Gueutin, Solenn Cordonnier, Oscar Lima, Sophie Michon-Coudouel, Delphine Naquin, Julie Ferreira de Carvalho, Malika L. Aïnouche, Armel Salmon, Abdelkader Aïnouche

‘ORIGINAL ARTICLE’

The first complete chloroplast genome of the genistoid legume *Lupinus luteus*: Evidence for a novel major lineage-specific rearrangement and new insights on plastome evolution in the legume family

Authors

Guillaume E. Martin[1,2†], Mathieu Rousseau-Gueutin[1†], Solenn Cordonnier[1], Oscar Lima[1], Sophie Michon-Coudouel[3], Delphine Naquin[1,4], Julie Ferreira de Carvalho[1], Malika Aïnouche[1], Armel Salmon[1], Abdelkader Aïnouche[1*]

[1]. UMR CNRS 6553 Ecobio, OSUR (Observatoire des Sciences de l’Univers de Rennes), Université de Rennes 1 / Université Européenne de Bretagne, 35 042 Rennes, France

[2]. Current address: CIRAD (Centre de coopération Internationale en Recherche Agronomique pour le Développement), UMR AGAP, F-34398 Montpellier, France

[3]. Plate-forme Génomique Environnementale et Fonctionnelle, OSUR-CNRS, Université de Rennes 1, 35042 Rennes, France

[4]. Plate-Forme de Bioinformatique, Genouest INRIA/IRISA, Université de Rennes-1, 35042 Rennes, France

[5]. Current address: Plate-forme IMAGIF, FRC3115 CNRS, 91198 Gif sur Yvette Cedex, France

[†]. The two first authors contributed equally

Short title: Novel lineage-specific inversion and insights in legume plastome evolution

*. Corresponding author: Dr. Abdelkader Aïnouche

email address: kader.ainouche@univ-rennes1.fr

Summary

Background and Aims To date chloroplast genomes are available only for members of the non-protein-amino-acid-accumulating clade (NPAAA) Papilionoid lineages in the legume family (i.e., Millettioids, Robinoids and the Inverted Repeat Lacking Clade). It is thus critical to sequence plastomes from other lineages to better understand the unusual evolution observed in this model flowering plant family. Therefore, the plastome of a lupine species, *Lupinus luteus*, was sequenced to represent the Genistoid lineage, one noteworthy but poorly studied legume group.

Methods The plastome of *L. luteus* was reconstructed using Roche-454 and Illumina next generation sequencing. Its structure, repetitive sequences, gene content and sequence divergence were compared to other Fabaceae plastomes. PCR screening and sequencing were performed in other allied legumes to determine the origin of a large inversion identified in *L. luteus*.

Key results The first sequenced Genistoid plastome (*Lupinus luteus*: 155,894 bp) allowed the discovery of a 36 kb inversion, embedded within the already known 50 kb inversion in the Large Single Copy region of the Papilionoideae. This inversion occurred at the base or soon after the Genistoid emergence, and most likely resulted from a flip-flop recombination between identical 29 bp inverted repeats within two *trnS* genes. Comparative analyses of chloroplast gene content of *L. luteus* versus Fabaceae and extra-Fabales plastomes revealed the loss of the plastid *rpl22* gene, and its functional relocation to the nucleus was verified using lupine transcriptomic data. Investigation on the evolutionary rate of coding and non-coding sequences among legume plastomes allowed the identification of remarkably variable regions.

Conclusions This study allowed the discovery of a novel major 36 kb inversion, specific to the Genistoids. We also identified chloroplast mutational hotspots, which contain novel potentially informative regions for molecular evolutionary studies at various taxonomic levels in legumes. Altogether, these results provide new insights on the legume plastome evolutionary landscape.

Key words: *Lupinus luteus*, chloroplast genome evolution, structural plastid rearrangement, 36 kb inversion associated with inverted repeats, flip-flop recombination, lineage-specific marker, functional gene transfer, Papilionoidae, repeated plastid sequences, sequence divergence, plastome hotspots, Fabaceae phylogeny.

INTRODUCTION

Legumes (Fabaceae) are the third largest angiosperm family with 727 genera and about 20,000 species (Lewis *et al.,* 2005). They are characterized by a wide biological and ecological diversity (Cronk *et al.*, 2006), and they are of great economic importance, particularly for human consumption or as animal forage. This family is composed of two main groups (Fig. 1): Caesalpinioideae *sensu lato* (including Mimosoideae) and Papilionoideae (Wojciechowski *et al.*, 2004; Cardoso *et al.*, 2012). The Papilionoideae is divided into six major clades: the Genistoids, Dalbergioids, Mirbelioids, Millettioids, Robinioids and the Inverted Repeat Lacking Clade (IRLC) (Cronk *et al.*, 2006). Within the Genistoids, *Lupinus* displays particular functional properties compared with other legumes, such as active nitrogen metabolism and the production of allelopathic substances of ecological and agronomical interest (Guillon and Champ, 2002; Magni *et al.*, 2004; Pilvi *et al.*, 2006). Additionally, *Lupinus* is the only known legume that does not form mycorrhizal

symbioses (Sprent, 2007). They also have considerable potential for phytoremediation due to their ability to metabolize nitrogen pollutants such as atrazine (Garcinuno *et al.*, 2003).

Plastome organization is highly conserved among most flowering plants (Jansen and Ruhlman, 2012), most having a quadripartite structure composed of two copies of an inverted repeat (IR) separated by large and small single copy regions (LSC and SSC). However, a few angiosperm families, including the Fabaceae, present an unusual plastome structure and evolution. In this family, the loss of one IR in the Papilionoideae (Wojciechowski *et al.*, 2004), the presence of many repetitive sequences (Saski *et al.*, 2005; Magee *et al.*, 2010), the occurrence of relatively large inversions (Palmer and Herbon, 1988; Perry *et al.*, 2002; Magee *et al.*, 2010) and the presence of a localized hypermutable region (Magee *et al.*, 2010) has been detected. Aberrant DNA repair was inferred as a possible cause for these plastomic rearrangements and accelerated rates of nucleotide substitutions (Jansen *et al.*, 2007).

Although most photosynthetic angiosperm plastomes contain 79 protein-coding genes, various derived lineages exhibit slightly fewer genes. Most of these rare chloroplastic gene losses occurred in species whose plastomes are highly rearranged relative to the ancestral angiosperm plastome (Jansen *et al.*, 2007). Since the emergence of the Fabaceae, there has been loss of five different chloroplastic genes: *accD*, *psaI*, *rpl23*, *rps16* and *ycf4* (Jansen *et al.*, 2007; Magee *et al.*, 2010). It is very likely that the genes lost from the plastome were previously functionally transferred to the nucleus or replaced by a nuclear gene of prokaryotic or eukaryotic origin. For example, the chloroplast *accD* gene was functionally transferred to the nucleus in *Trifolium* species (Magee *et al.*, 2010) and the plastidic *rps16* gene was functionally replaced by a nuclear-encoded *rps16* gene of mitochondrial origin in *Medicago truncatula* (Ueda *et al.*, 2008).

During the last decade, knowledge on the organization and evolution of legume plastomes has rapidly expanded with the development of next generation sequencing technologies (NGS).

Ten legume plastomes have now been sequenced: *Cicer arietinum* (Jansen *et al.*, 2008), *Glycine max* (Saski *et al.*, 2005), *Lathyrus sativus* (Magee *et al.*, 2010), *Lotus corniculatus* (Kato *et al.*, 2000), *Medicago truncatula* (unpublished data), *Millettia pinnata* (Kazakoff *et al.*, 2012), *Phaseolus vulgaris* (Guo *et al.*, 2007), *Pisum sativum* (Magee *et al.*, 2010), *Trifolium subterraneum* (Cai *et al.*, 2008), and *Vigna radiata* (Tangphatsornruang *et al.*, 2010). The sequencing of these plastomes confirmed previous observations of major rearrangements in this family, including a 50 kb inversion present in most papilionoids (Palmer and Thompson 1982; Lavin *et al.*, 1990; Doyle *et al.*, 1996; Wojciechowski *et al.*, 2004; Jansen *et al.*, 2008) and the loss of one copy of the IR region in one of the papilionoid clade, called the *Inverted Repeat Lacking Clade* (IRLC) (Palmer and Thompson, 1982; Lavin *et al*., 1990; Wojciechowski *et al.*, 2004; Jansen *et al.*, 2008). However, all the Papilionoideae plastomes sequenced to date belong to three clades (Millettioids, Robinioids and IRLC) within the non-protein-amino-acid-accumulating clade (NPAAA clade; according to Cardoso *et al.*, 2012). Thus it is essential to investigate representatives from other Papilionoid lineages to better understand plastome evolution within the Papilionoideae, and more broadly within legumes. In this context, the genus *Lupinus* is a good candidate to represent the core Genistoids (Wojciechowski *et al.*, 2004; Cronk *et al.*, 2006; Cardoso *et al.*, 2012) that is one of the poorly studied legume lineages. Although considerable strides have been made in elucidating the evolutionary history of the Fabaceae using plastid DNA sequence-based phylogenies (Wojciechowski *et al.*, 2004; Cardoso *et al.*, 2012), there is still a great need to more accurately elucidate relationships at other taxonomic levels among and within lineages of the 50 kb inversion Papilionoid clade, including within the Genistoids and in the genus *Lupinus* (Ainouche and Bayer, 1999; Ainouche *et al.*, 2004; Hughes and Eastwood, 2006; Drummond, 2008; Mahé *et al.*, 2011a; Mahé *et al.*, 2011b). Therefore, the lupine plastome sequence not only provides the raw material to extend understanding of legume genome

organization and evolution, but also provides an important source of phylogenetically informative plastid molecular markers, which have the advantage of being uniparentally (maternally) inherited and generally non-recombinant (Jansen *et al.*, 2007; Moore *et al.*, 2007; Moore *et al.*, 2010) .

In this paper, we report the complete sequence of the chloroplast genome of *Lupinus luteus*, the first sequenced in the core Genistoids. After reconstruction and annotation, this genome has been compared with other Fabaceae and extra-Fabales plastomes, allowing the identification of a noteworthy 36 kb inversion. A PCR and sequencing survey of this inversion across various legume representatives provided evidence that this inversion represents a novel genomic rearrangement, characterizing the core Genistoids. The gene content within the *L. luteus* plastome has also been compared to other Fabaceae and closely related species in order to identify chloroplast genes lost from the *L. luteus* plastome. We verified that the chloroplast genes missing in the *Lupinus* plastome were functionally transferred to the nucleus. Finally, we evaluated the sequence divergence between the lupine and other Fabaceae plastomes at different levels (exon, intron, and intergenic) in order to better understand the unusual plastome evolution and to suggest potentially useful plastid regions for molecular phylogenetic analyses in Fabaceae.

MATERIALS AND METHODS

DNA extraction, high-throughput sequencing and isolation of chloroplast sequences

Genomic DNA was extracted from fresh leaves of an individual sample (Lab. collection ref. number: M6=EGSM6Llu2) from a natural population of *Lupinus luteus* collected at Bou Tlelis, Oran in Algeria/North Africa. DNA extraction was performed using NucleoSpin® Plant II kit (Macherey Nagel) following the manufacturer's instructions. The genomic DNA was subjected to two high-throughput methods of sequencing: one run using pyrosequencing

with the GS-FLX (454 Life science – ROCHE) platform (OSUR/biogenouest; Université de Rennes-1) that generated 799,732 reads of ca. 400 bases, and one flow cell lane performed with an Illumina HiSeq 2000 platform (BGI, Hong Kong) that yielded 11.46 million of 2*100 bases paired-end reads from a library of ca. 500 base DNA fragments. Reads corresponding to plastome sequences were extracted from the Roche-454 dataset using blast similarity search (e-value $10^{-6,}$ 90% identity) against the fully sequenced plastomes of *Glycine max* (NC_007942), *Medicago truncatula* (NC_003119), *Lotus japonicus* (NC_002694), *Cicer arietinum* (NC_011163), *Pisum sativum* (NC_0147057), *Trifolium subterraneum* (NC_011828), *Lathyrus sativus* (NC_014063), *Millettia pinnata* (NC_016708), *Vigna radiata* (NC_013843), *Phaseolus vulgaris* (NC_009259), *Populus trichocarpa* (NC_009143) and *Arabidopsis thaliana* (NC_000932). A total of 21,018 reads corresponding to plastid sequences were obtained from the 454 sequencing and 509,962 paired-end reads from Illumina.

Plastome assembly and annotation

De novo assembly was performed from filtered Roche-454 reads using Newbler (v. 2.5.3, 454 Life Science). A total of 45 contigs ranging from 450 to 25,000 bases were obtained and organized using the *Glycine max* plastome as a reference. Illumina paired-end reads having at least one mate mapping with Bowtie (Langmead *et al.*, 2009) on the 45 contigs were extracted from Illumina dataset. The draft plastome sequence as well as the junctions between contigs were verified and corrected with the 509,962 paired-end Illumina reads extracted using Mira v. 3.4.0 (Chevreux *et al.*, 1999) and Bowtie (Langmead *et al.*, 2009). The 454 and Illumina datasets allowed a 73x (standard deviation: 53) and 884x coverage (standard deviation: 466) of the newly reconstructed *L. luteus* plastome, respectively.

Plastome annotation was conducted in four steps: 1) Identification of protein encoding sequences by aligning (blastp, e-value threshold: 10^{-5}) *Glycine max* protein-coding sequences obtained from ChloroplastDB (Cui *et al.*, 2006) against chloroplastic ORFs extracted from the *Lupinus* plastome sequence using the perl script ***get_orf.pl*** designed by Paul Stothard (University of Alberta); 2) Identification of ribosomal RNA (rRNA) and transfer RNA (tRNA) sequences by direct alignment of *Glycine max* tRNAs and rRNAs against the *Lupinus* plastome sequence; 3) Verification of the identification of all plastomic genes using DOGMA (Wyman *et al.*, 2004); 4) Verification of the annotation by performing manual alignment using BioLign and multiple contig editor (v. 4.0.6.2). A graphical representation of the chloroplast genome was performed using the CIRCOS software (Krzywinski *et al.*, 2009).

To determine the presence of codon bias, the number of codons ending with A-T or C-G was tallied and a Chi square test was performed for each amino acid. These tests were subjected to a Bonferroni correction for multiple testing performed with the R software package (http://www.r-project.org).

Identification of repeat elements

The number and location of repeated elements (tandem, palindrome, dispersed direct and dispersed inverted repeats) in *Lupinus luteus* plastome were determined using REPuter (Kurtz *et al.*, 2001). We used the same parameters than previously used for other Fabaceae species (Saski *et al.*, 2005; Cai *et al.*, 2008; Tangphatsornruang *et al.*, 2010). More precisely, we searched for repeated elements with a minimum size of 30 bp and a Hamming distance of 3 (sequence identity of ≥90%). One copy of the inverted repeat (IR) was removed before performing the analysis.

Identification and origin of the 36 kb inversion by PCR screening and sequencing

In order to identify the putative presence of large structural variation (> 1 kb) within the *L. luteus* plastome, breaks of synteny were searched between plastomes of *L. luteus*, other legumes and two outgroup taxa (*Cucumis sativus* from the Cucurbitales and *Prunus persica* from the Rosales) by performing dot plots using the Gepard software (Krumsiek *et al.*, 2007).

To determine the origin of the large inversion observed in *Lupinus luteus*, its presence/absence was surveyed by PCR in *Lupinus* and in representatives of various genera more or less closely related to the lupines in the core Genistoids: *Argyrolobium uniflorum*, *Chamaecytisus mollis*, *Crotalaria saharae, Echinospartum boissieri*, *Genista florida*, *Genista tricuspidata*, *Laburnum anagyroides*, *Lupinus microcarpus*, *Retama sphaerocarpa*, *Sophora japonica*, *Thermopsis rhombifolia*, *Ulex minor*. Outgroup taxa were also screened for the presence/absence of this inversion, for instance: *Cercis siliquastrum* that is basal in the legume family; *Acacia dealbata* that belongs to the Mimosoids; and *Cladrastis lutea*, a Papilionoid that is sister to the 50 kb inversion clade. A PCR strategy using primer pairs diagnostic for the presence or absence of the inversion was conducted. The primer pairs were designed in either conserved *ycf3* and *psbI*, or *rps4* and *ycf3* protein-coding sequences, which are flanking the inversion end points, to allow the assessment of the presence or absence of the inversion.

Each PCR amplification was performed in a total volume of 50 μL containing 10 μL of 5× Go taq green buffer (Promega), 5 μL of 2 mM deoxyribonucleotide mix, 4 μL of each primer (5 mM), 0.2 μL of Go Taq polymerase (5 units/μL) and 20 ng of template DNA. Cycling conditions were 94°C for 2 min, followed by 35 cycles of 94°C for 45 s, 55°C for 30 s, and 72°C for 90 s, and a final extension of 72°C for 7 min. The primer pairs used to detect the absence or presence of the 36 kb inversion were: rps4-bef-F (5'-CAATCAAATAATAGATAGTAAATGGGTTG-3') and ycf3-bef-R (5'-GGAATTATTCGTAATAATATATTGGCTAC-3'); ycf3-inv-F (5'-

CGTAATAAGATATTGGCTAC-3') and psbI-int-R (5'-CTCTTTTCATCTTCGGATTC-3'). The PCR products were then purified using the NucleoSpin Gel and PCR Clean-up purification kit (Macherey-Nagel) and sequenced directly in both directions (Macrogen Europe, Amsterdam).

Evolution of the gene content in Fabaceae plastome and identification of genes functionally transferred to the nucleus in *Lupinus*

In order to determine whether *L. luteus* has recently lost chloroplastic genes, its plastome was compared to those of ten other legume species (*Millettia pinnata*, *Vigna radiata*, *Glycine max*, *Phaseolus vulgaris*, *Trifolium subterraneum*, *Medicago truncatula*, *Lotus japonicas*, *Cicer arietinum*, *Pisum sativum* and *Lathyrus sativus*) and two ougroup species available in GenBank. During Fabaceae evolution, five chloroplastic genes (*accD*, *psaI*, *rpl22*, *rpl23* and *rps16*) have been lost from the plastome of various lineages, of which three (*accD*, *rpl22* and *rps16*) were shown to have been independently functionally relocated to the nucleus or replaced by a nuclear gene in different Fabaceae (Gantt *et al.*, 1991; Millen *et al.*, 2001; Ueda *et al.*, 2008; Magee *et al.*, 2010). We searched for putative functional transfer to the nucleus (functional relocation or intermediate stage) of these five plastid genes within transcriptomic data available from our laboratory for *Lupinus mariae josephi* (unpublished data). The identification of these putative functional transfers was performed by blasting (e-value threshold: 10^{-10}) the following sequences against the transcripts of *L. mariae josephi*: the *Trifolium repens* nuclear *accD* (Magee *et al.*, 2010) and the *L. luteus* plastidic *accD* genes; the *Pisum sativum* nuclear *rpl22* sequence (Gantt *et al.*, 1991); the plastidic *psaI*, *rpl23* and *ycf4* genes from various Fabaceae (*Lupinus luteus*, *Lotus japonicus* and *Phaseolus vulgaris*); the *Medicago truncatula* nuclear-encoded *rps16* genes of mitochondrial origin (Ueda *et al.*, 2008). The presence of a transit peptide-encoding sequence within the identified chloroplastic genes functionally replaced in the nucleus were then predicted using BaCelLo (Pierleoni *et*

al., 2006), Predotar (Small *et al.*, 2004) and TargetP (Emanuelsson *et al.*, 2000) software programs. To confirm that the nuclear *rpl22* gene identified in *Lupinus mariae josephi* results from an early functional transfer to the nucleus in the common ancestor of all flowering plants, as demonstrated with *Pisum sativum* by Gantt *et al*. (1991), we aligned these sequences (*Lupinus* and *Pisum*) with the *rpl22* amino-acids sequences from eubacteria, algae, bryophytes and land plants using Geneious package (http://www.geneious.com/). After excluding the extreme 5' and 3' ends of the sequences, a data matrix of 98 amino-acids was subjected to phylogenetic analyses using PHYML (Guindon and Gascuel, 2003) and Neighbor-Joining (Saitou and Nei, 1987). The tree was rooted using the eubacteria *Mycoplasma*. Bootstrap values were performed with 1,000 replicates (Felsenstein, 1985).

Evaluating sequence divergence between the complete lupine plastid genome and those from other legumes and Fabids

Sequence divergence between *L. luteus* and ten other Fabaceae plastomes was evaluated independently for each homologous regions aligned with MUSCLE (Edgar, 2004). Pairwise distances were calculated with the ***ape*** R-cran Package (Paradis *et al.*, 2011, available at http://cran.r-project.org/web/packages/ape/ape.pdf) using the Kimura-2-parameters (K2p) evolution model (Kimura, 1980). The mean sequence divergence rate of the different genetic categories (*i.e.* IGS, intron, rRNA and tRNA, exons) was compared using Mann-Whitney test with Bonferroni correction. Additionally, estimation of sequence divergence of coding-protein genes (exons) were calculated using the synonymous (Ks) and non-synonymous (Ka) nucleotide substitution rates with the yn00 method (Yang and Nielsen, 2000) from the PAML package (Yang, 2007). Finally, fast evolving sequences were identified. Only the protein-coding, intronic or intergenic regions presenting a higher evolutionary rate than the regions most commonly used for evolutionary studies in Fabaceae (*rbcL* and *matK* genes, the 5'*trnK*

and *trnL*-introns, and the *trnK_trnF*, *trnL_trnT* and *trnS_trnG* intergenic spacers) and a minimum size of about 300 bp were retained.

A list of all the softwares used in this study, their purpose and availability can be found in the Supplementary Data Table S1.

RESULTS

Organisation, gene content and characteristics of the *L. luteus* plastome

The *Lupinus luteus* plastome (deposited in GenBank: KC695666) has a length of 151,894 base pairs (bp), with a quadripartite structure composed of two IRs (25,860 bp) separated by a SSC (17,847 bp) and a LSC (82,327 bp) region (Fig. 2). It contains 111 different genes, including 77 protein-coding genes, 30 transfer RNA (tRNA) genes and four ribosomal RNA (rRNA) genes (Table 1). Protein-coding genes, tRNA and rRNA represent respectively 51.6%, 1.8% and 6.0% of the plastome. Non-coding DNA, including intergenic spacers (IGS) and introns represent 40.6% of the genome. The overall GC content of *L. luteus* plastome is of 36.6%. It is higher in tRNA and rRNA (53.3% and 55.3% respectively), slightly higher in protein-coding genes (37.3%), similar in introns (36.3%) and lower in IGS (30.3%).

The *Lupinus luteus* plastome contains 18 different intron-containing genes (of which six are tRNA), as in most Fabaceae species. All intronic genes contain one intron, apart from two genes (*clpP* and *ycf3*) that contain two introns. Within the IR, four rRNA, seven tRNA and five protein-coding genes are repeated. Only the 5' end of the *ycf1* gene (519 bp) is present in the IR and the gene *rps12* is trans-spliced, with the 5'exon in the LSC and the remaining two exons in the IR.

Thirty different tRNA are present in *L. luteus* plastome. They correspond to 28 different codons, at least one for each amino acid. Seven of the 28 different anticodon tRNAs encoded in the *Lupinus* plastome correspond to the most common codon (where synonymous codons

exist). The codon usage is biased towards a high representation of A and T at the third position (Supplementary Data Table S2).

Repeat elements in *Lupinus luteus* plastome

All repeat sequences that present a minimum size of 30 bp and with a sequence identity ≥90% were identified in *Lupinus luteus* plastome using REPuter (Kurtz *et al.*, 2001). A total of 31 repeats were found (Supplementary Data Table S3), including 13 dispersed inverted repeats, nine palindromes, six dispersed direct and three tandem repeats.

All the palindromic repeats observed in *L. luteus* plastome are localized in intergenic regions (except one in *ycf1* coding sequence) and tandem repeats are mainly found in coding sequences (*ycf2*).

Most repeats (93,55%) are 30-50 bp long. The largest repeat in the plastome is a 288 bp dispersed direct repeat corresponding to a fragment of *ycf2* duplicated in each IR, between *rpl23* and *tnrI_CAU*. This repeated element is absent from extra-Fabales plastomes (*Cucumis sativus*, *Prunus persica*) or from IRLC plastomes in the Papilionoideae, but is present in other non IRLC Papilionoideae plastomes (*Glycine max*, *Lotus japonicus*, *Phaseolus vulgaris*, *Vigna radiata*), as previously observed by Guo *et al*. (2007).

Origin of a 36 kb inversion detected in *Lupinus luteus*

Global alignment and comparison of gene order between the plastomes of *L. luteus* and other legumes, as well as with outgroups, revealed an inversion of about 36 kb between the *trnS-GCU* and *trnS-GGA* genes in *L. luteus* (Fig. 3). This inversion is highlighted by Dot plot analyses that compared the plastomes of *L. luteus* with either *Glycine max* or *Cucumis sativus* (Supplementary Data Fig. S1). This unique 36 kb inversion is embedded within the 50 kb

inversion that occurred earlier in the Papilionoideae after the divergence between the *Cladrastis* clade and the rest of the more derived papilionoid legumes (Doyle *et al.*, 1996).

To verify the existence of this inversion in *Lupinus* and to screen other Genistoids and legume species for the presence or absence of this 36 kb inversion, two diagnostic primer pairs were designed. The localization of these primers is indicated in Fig. 3. PCR amplification was expected from the primers located in the *rps4* and *ycf3* protein-coding sequences only for the species without the inversion, whereas PCR amplification using primers within *ycf3* and *psbI* genes was only expected in species with the inversion. Since the 36 kb inversion identified in *L. luteus* is not present in the plastomes available for extra Fabaceae taxa or in the derived Fabaceae (representative of the Millettoid, Robinoid and IRLC clades), it most likely occurred after the emergence of the Genistoids. For the 11 core Genistoid species screened here (including representatives of the Sophoreae, Thermopsidae and Genisteae tribes), amplifications were only successful when using the diagnostic primers pair for the presence of the inversion. In contrast, all non-core Genistoids tested gave amplification only when using the diagnostic primers pair for the absence of the 36 kb inversion (Fig. 4). In both cases the results were confirmed by sequencing of the PCR products (deposited in Genbank: KC695667 - KC695681) and alignment of the sequences with the homologous plastidic region from other Fabaceae plastomes, including *Lupinus luteus*.

Several multiple alignments of the sequences surrounding the endpoints of the 36 kb inversion using Papilionoideae species with or without the 36 kb inversion allowed determining the exact location of the inversion (Supplementary Data Fig. S2). It occurred between the 3'end of the *trnS-GGA* and the *trnS-GCU* that are identical for the last 29 bp and are in inverse orientation (Fig. 5; Supplementary Data Table S3). A similar inverted repeat also exists between *trnS-GGA* and *trnS-UGA* that are 9 kb distant.

Gene content and gene transfers to the nucleus in *Lupinus* compared to other Fabaceae

The protein-coding gene content of *Lupinus luteus* plastome was compared to those of ten other Fabaceae and two outgroup taxa. The aim of this comparison was to examine whether the lupines (representing the core Genistoids) have lost or retained the chloroplast genes known to have lost their functionality in the plastome of various lineages during legume evolution (reviewed in Magee *et al.* 2010) such as: *accD*, *psaI*, *rpl22*, *rpl23*, *rps16* and *ycf4* (as indicated in Fig. 6). Out of these six plastidic genes lost from legume lineages, only *rpl22* is missing in the plastome of *L. luteus*. The functional transfer of this gene to the nucleus, already demonstrated in *Pisum sativum* (Gantt *et al.*, 1991), was verified in a lupine species (*Lupinus mariae josephi*) by the identification of a nuclear *rpl22* transcript that is similar to the nuclear *rpl22* transcript found in *Pisum sativum*. The presence of a chloroplast target peptide was predicted in the *L. mariae josephi* nuclear *rpl22* transcript using a variety of software (data not shown). The alignment and phylogenetic analysis of nuclear and chloroplastic *rpl22* sequences (Supplementary Data Fig. S3) showed that the nuclear *rpl22* gene observed in *Lupinus* and *Pisum* derive from the same transfer event, which occurred in the common ancestor of all flowering plants (Gantt *et al.*, 1991). Concerning the other chloroplast genes lost during Fabaceae evolution, investigations were performed to determine whether they could be at an intermediate stage of functional transfer to the nucleus. We identified nuclear *rps16* transcripts in *Lupinus mariae josephi* that were similar to the *M. truncatula* nuclear *rps16* genes (Ueda *et al.*, 2008) but no nuclear *accD*, *psaI*, *rpl23* or *ycf4* transcripts could be detected.

Sequence divergence between the plastome of *L. luteus* and other Fabaceae

A comparison of pairwise distances (K2p) calculated for non-coding regions between *L. luteus* and other legumes (Supplementary Data Table S4) revealed that, as expected, IGSs

evolve significantly more rapidly than introns. Slowest evolving regions are tRNAs and rRNAs (Supplementary Data Table S4). For introns (Fig. 7A), the mean sequence divergence ranged from 0.028 (for *rps*12 intron) to 0.270 (for *clpP* intron1). The two main introns previously used for phylogenetic inference in legumes showed relatively low rates of variation: 0.100 for the *trnL* intron (501 bp length) and 0.148 for the *trnK* 5' intron (318 bp length). Among introns, seven exhibited a higher level of divergence (Fig. 7A), such as: *trnG-UCC* intron (K2p = 0.183; 698 bp), *rpoC1* intron (0.164; 766 bp), *clpP* intron 2 (0.230; 739 bp), *clpP* intron 1 (0.270; 655 bp), *petD* intron (0.224; 743 bp), *rpl16* intron (0.195; 1155 bp), and *ndhA* intron 0.215; 1171 bp). The highest mean sequence divergence of IGS regions corresponds to the *accD_psaI* region (0.473). Among IGS and comparatively to the most used IGS regions for legumes phylogeny (*trnL_trnF*, mean K2p = 0.255; *trnL_trnT*, 0.352; *trnS_trnG*, 0.316), five IGS larger than 300 bp showed divergence rates slightly or significantly higher than *trnL_trnT*: *ycf4_cemA* (0.357; 317 bp); *rpl36_rps8* (0.357; 453); *psbZ_trnG-GCC* (0.357; 345 bp); *trnV-UAC_ndhC* (0.355; 497 bp), and *accD_psaI* (0.473; 293 bp (Fig. 7A). For protein-coding regions (Supplementary Data Table S5; Fig. 7B), the evolutionary rates have been evaluated by comparison of their synonymous (Ks) nucleotide substitution rates (Fig. 7B). The mean divergence rate between *L. luteus* and the other legume genes ranged fom 0.072 (for *rpl23*; 282 bp) to 0.667 (for *rps16*; 47 bp), with most loci presenting mean Ks values lower than those of the two protein-coding genes used for phylogenetic inference in legumes, *matK* (0.235; 1521 bp) and *rbcL* (0.367; 1428 bp). Fourteen genes displayed Ks values higher than *rbcL*, of which nine are larger than 300 bp: *rpoC2* (0.419; 4149 bp); *rps16* (0.667; 407 bp); *accD* (0.538; 1497 bp); *ycf4* (0.659; 555 bp), *rps8* (0.388; 405 bp), *rpl14* (0.441; 369 bp), *ycf1* (0.518; 5296 bp), *ndhH* (0.394; 1182), *ndhF* (0.454; 2241 bp). For most loci (65/77), the nonsynonymous nucleotide substitutions (Ka) values calculated between *Lupinus* and the other legumes were lower than 0.1

(Supplementary Data Table S6). Among the twelve remaining loci, only five displayed higher values than the reference *matK* gene (mean Ka = 0.132): *rpl32* (0.147), *rps16* (0.169), *accD* (0.184), and particularly *ycf1* (0.306) and *ycf4* (0.398). Regarding these low values of Ka, the *Ka/Ks* ratio calculated for each protein-coding gene (Supplementary Data Table S7) was lower than 1 and even lowest than 0.5 for almost all loci, indicating that plastidic genes are under a high negative (*i.e.,* purifying) selective constraint (Kimura, 1977; Messier and Stewart, 1997). Nevertheless, it can be noticed that the three *ycf* genes (*ycf1*, *ycf2* and *ycf4*) exhibit remarkably higher *Ka/Ks* values (0.601, 0.682 and 0.649, respectively) than all the other genes, indicating an increase of their sequence evolutionary rate.

Altogether, these analyses allow circumscription of fast evolving regions in the legume plastomes, as inferred from pairwise comparison of *Lupinus* to the other available legume plastomes (highlighted in Supplementary Data Fig. S4). Among these regions, three are remarkable: one in the SSC, between *ycf1* and *ndhA* intron (9043 bp); two in the LSC, around the *rpl36-rpl16* genes (3178 bp), and the *accD-ycf4_cemA* region (2968 bp) that exhibits the highest rates of sequence divergence for genes (*rps16, accD* and *ycf4*) and IGS (*accD_psaI*, and *ycf4_cemA*). The latter region, which includes the *ycf4* gene, was shown to have a dramatic increase of its evolutionaty rate in the NPAAA clade (incl. Millettioids, Robinioids, and IRLC) and most particularly in *Lathyrus* (Magee *et al.*, 2010). To investigate whether such acceleration also occurred in the lupine lineage, maximum likelihood phylogenetic analyses using legume *ycf4* gene sequences (including *L. luteus ycf4*) and based on Ks and Ka substitution rates were performed. Our results (Supplementary Data Fig. S5) do not provide evidence of such acceleration in the *ycf4* gene in *Lupinus*, in accordance with the previous results obtained by Magee *et al.* (2010) using a few Genistoid members (*Crotalaria, Goodia, Laburnum*).

The other regions showing peaks of divergence when comparing *Lupinus* with other legume plastomes include some isolated genes (*rpoC2* and *ndhF*), introns (*trnG-UCC, rpoC1*) and *petD*) and IGS (*trnV_ndhC*; *psbZ_trnG-GCC*) that were not previously detected as fast evolving regions in NPAAA clade members.

DISCUSSION

In this work the plastome of *Lupinus luteus* has been sequenced using Next Generation Sequencing technologies. Its size and gene content are within the range found in plastomes containing two IR (Raubeson and Jansen, 2005). It is AT rich (exception: rRNA and tRNA genes) and the codon usage is biased toward a high representation of A and T at the third position as previously observed by Clegg *et al.* (1994). This sequence, which represents the first plastome sequenced in the core-Genistoids, is of major interest because all legume plastomes sequenced so far belong to only three Papilionoid clades, the Millettioids, the Robinioids and the IRLC, which derive from within the non-protein-amino-acid accumulating (NPAAA) clade (Cardoso *et al.*, 2012). Thus it was essential to sequence plastomes from representatives of other Papilionoid lineages in order to have a better understanding of the unusual plastome evolution observed in legumes (Jansen and Ruhlman, 2012). Most photosynthetic angiosperms have a highly conserved plastome organization, except the Campanulaceae, Fabaceae and Geraniaceae families that exhibit remarkable and extensive rearrangements (Jansen and Ruhlman, 2012). Within the Fabaceae, one of the most remarkable inversion that occurred after the emergence of the family is the 50 kb inversion, which is shared by most papilionoid taxa (Doyle et al., 1996). The plastome sequences of three IRLC species, *Lathyrus sativus* (Magee *et al.*, 2010), *Pisum sativum* (Palmer and Herbon, 1988) and *Trifolium subterraneum* (Cai *et al.*, 2008), revealed that, relative to the ancestral angiosperm plastome organization, and after the 50 kb inversion event, six, eight

and fifteen inversions occurred, respectively (Magee *et al.*, 2010). Within the 50 kb clade, our study reveals that the *L. luteus* plastome experienced an additional 36 kb inversion internal to the 50 kb inversion, which most likely occured at the origin of the core Genistoids. Previous molecular characterization of large plastomic inversion endpoints in a few plant families or genera, including the 50 kb inversion present in most Papilionoideae (Doyle *et al.*, 1996), the 22 kb inversion in Asteraceae (Kim *et al.*, 2005), the 42 kb inversion in *Abies* (Tsumura *et al.*, 2000) or the 21 kb inversion in Jasminae (Lee *et al.*, 2007), showed that these large plastomic inversions were often associated with short inverted repeats present within, or adjacent to, a tRNA. The detailed survey of the regions surrounding the 36 kb inversion endpoints in core-Genistoids allowed us to determine that this inversion is most likely due to the presence of inverted repeat motifs (29 identical nucleotides) at the 3'end of the *trnS-GGA* and *trnS-GCU*. The role of repeated elements present in inverse orientation in promoting flip-flop recombination resulting in inversions has been previously demonstrated using tobacco transplastomic lines (Rogalski *et al.*, 2006). Such repeated elements can promote plastid DNA inversions which may vary in size from a few base pairs to several kilobases (reviewed in Downie and Palmer, 1992). Minor inversions are more common than major ones and mainly occur in noncoding regions, intergenic spacers and introns (Palmer, 1995). Interestingly the inverted repeat motif in the *trnS-GGA* and *trnS-GCU*, which caused the 36 kb inversion in the core-Genistoids, is present in almost all Rosids (Supplementary Data Table S8) and is separated by at least 30 kb. Thus, even though this 36 kb inversion was only observed in the core-Genistoids, it could have occurred in any other rosid species.

The *Lupinus luteus* plastome contains fewer repeats (31) than other Fabaceae species, such as *Vigna radiata*, *Lotus japonicus*, *Glycine max* or *Medicago truncatula* that have 50, 67, 104 and 191 repeats, respectively (Saski *et al.*, 2005; Tangphatsornruang *et al.*, 2010). Most *L. luteus* repeats are relatively small in size (90 % are 30-50 bp in size). The longest repeat

observed in *L. luteus* is a 288 bp direct repeat (within the IR) that is also present in the non-IRLC Papilionidae (*Glycine max*, *Lotus japonicus*, *Phaseolus vulgaris* and *Vigna radiata*) but not in the IRLC or outgroup taxa (*Cucumis sativus* and *Prunus persica*). The low number of repeats observed in *Lupinus* is in stark constrast with *Trifolium subterraneum* plastome that contains a high number of large repeats and shows a high rate of rearrangement: 14 inversions occurred since its divergence with other IRLC species (Cai *et al.* 2008; Magee *et al.* 2010). The number of large repeats was demonstrated to be positively correlated to the degree of plastome rearragements in plants (Maul et *al.*, 2002; Pombert *et al.*, 2005; Guisinguer *et al.*, 2011). Within the repeats observed in *Lupinus*, 42 % (13/31) are dispersed inverted repeats (6 to 66 kb distant) that could promote inversions. However, apart from the dispersed inverted repeat at the origin of the 36 kb inversion in the core Genistoid and another repeat between the *trnS-UGA* and the *trnS-GGA* (9 kb distant), all the others would most presumably lead to the loss of functionality of one or several genes in case of an inversion event and thus may be deleterious (Ruf *et al.,* 1997: Drescher *et al.,* 2000). Whilst the above 9 kb region is potentially prone to inversion, to date no evidence of such event has been observed from this or previous studies in other Fabaceae species.

The rarity of plastomic rearrangements in flowering plants makes these characters powerful phylogenetic markers (Kim *et al.*, 2005) since they present an extremely low level of homoplasy (Cosner *et al.*, 2004). The 36 kb inversion identified in this study is present in all core Genistoid species surveyed (12) and therefore provides a robust additional synapomorphy supporting monophyly of the core Genistoids (Crisp *et al.*, 2000). Further screening of representatives from Brongniartieae and Bowdichia clades, shown to be closely related to the core Genistoids (Cardoso *et al.*, 2012), will determine whether this 36 kb inversion is strictly specific to the core Genistoids or whether it occurred earlier or at the base of the large Genistoid *s.l.* assemblage (includes Brongniartieae and Bowdichia clades). Thus,

after the 50 kb inversion that is shared by a majority of Papilionideae (Doyle *et al.*, 1996), and the 78 kb inversion supporting the Papilionoid subtribe Phaseolinae (Bruneau *et al.*, 1990) in legumes, this 36 kb inversion represents an additional example highlighting the phylogenetic usefulness of plastidic inversions. Such clade demarcating inversions were also detected in other Angiosperm families. Within the Asteraceae, a 22 kb inversion allowed identification of the subtribe Barnadesiinae as the most primitive lineage in the family (Jansen and Palmer, 1987). In the Campanulaceae, which also have highly rearranged plastomes, reliable phylogenetic relationships could be reconstructed within the family only based on the use of the numerous rearrangements (including inversions) as characters (Cosner *et al.*, 2004). Interestingly, there is also evidence of specific mutational and restructuring events that affected the nuclear genome of *Lupinus* (Mahé *et al.*, 2011a), which suggests that the Genistoids experienced noteworthy genomic changes, in both the plastid and the nuclear genomes, after their divergence from the NPAAA papilionoid lineages (~50-56 My ago).

Gene content is highly conserved among photosynthetic angiosperm plastomes (Timmis *et al.*, 2004). However, within the Fabaceae, several chloroplastic genes (*accD*, *psaI*, *rpl22*, *rpl23*, *rps16*, *ycf4*) have been lost recently and independently in various lineages (Magee *et al.*, 2010). However, within the *Lupinus luteus* plastome, only the *rpl22* gene is missing, which is in accordance with the previous finding of Gantt *et al.* (1991) who demonstrated that the functional transfer of this gene from the chloroplast to the nucleus occurred in a common ancestor of all flowering plants, and thus preceded its loss from the chloroplast genome by about 100 million years (Supplementary Data Fig. S3). Among the chloroplast genes that have been lost in legume lineages following their divergence from the common ancestor with lupines, we found that the *rps16* gene is at an intermediate stage of functional replacement in the *Lupinus* nuclear genome, as it is still represented by a functional copy in the chloroplast genome while another is in the nucleus. This nuclear-encoded *rps16* gene targeted to the

plastid is of mitochondrial origin and was transferred prior to the monocot-dicot divergence (Ueda *et al.*, 2008).

The evaluation of sequence divergence between *Lupinus* and the other sequenced legumes allowed identification of fast-evolving sequences (Fig. 7; Supplementary Data Fig. S4). This information is essential to better understanding the dynamic nature of plastome evolution in legumes and to improving legume phylogeny, especially within the Genistoids and the genus *Lupinus* (Eastwood *et al.*, 2008; Mahé *et al.*, 2011a; Cardoso *et al.*, 2012). As expected, most coding regions are well conserved, particularly in the IR region, and in most cases IGS are evolving faster than introns, in accordance with previous observations (Clegg and Zurawski, 1991; Raubeson *et al.*, 2007). Compared to the plastid sequences used in legume evolutionary studies, we have detected several sequences (Fig. 7) that exhibit higher rates of divergence: (i) seven introns (*trnG-UCC*, *rpoC1*, *rpl16*, *ndhA*, *petD*, *clpP* introns 1 and 2) (ii) five intergenic spacers (*ycf4_cemA*, *rpl36_rps8*, *psbZ_trnG-GCC*, *trnV-UAC_ndhC* and *accD_psaI*); and (iii) eight protein-coding genes (*rpoC2*, *accD*, *ycf4*, *rps8*, *rpl14*, *ycf1*, *ndhH* and *ndhF*). Interestingly, most of these variable regions have not been or rarely been employed in legumes phylogeny (*e.g.*: *trnS-trnG* region in *Lupinus* by Drummond, 2008; *ycf1* in *Astragalus* by Bartha *et al*. 2013), and thus represent a new set of markers to explore evolutionary relationships within legumes. Each of these sequences needs to be tested in order to evaluate at which taxonomic level and in which lineages they could be more informative and useful. As an example, the remarkable increase of the evolutionary rate observed in the *ycf4* gene is specific to the IRLC, Robinioid and Millettioid lineages (NPAAA clade), and occurred after the divergence of the latter from the other legumes (Magee *et al.* 2010). Thus, this region is most likely a good candidate for the NPAAA clade but seems less interesting for phylogenetic inference within the Genistoids and earlier legume lineages. In contrast, we have detected several loci exhibiting an increase of their evolutionary rate that is specific to the

lupine/genistoid lineage. These loci include the *rpoC2 and ndhF* genes, the *trnG-UCC*, *rpoC1* and *petD* introns, the *trnV_ndhC* and *psbZ_trnG-GCC* IGS. Regardless of the specificity and degree of utility of each locus, altogether these variable sequences constitute an important source of novel characters for single- or multigene-based reconstruction of evolutionary patterns in legumes at various taxonomic levels.

These variable sequences (mentioned above) are distributed in well-circumscribed fast evolving regions that shape the legume plastome landscape (Fig. 7; Supplementary data Fig. S4). Interestingly, three of these variable regions are located at boundaries of the 50 kb inversion (*rps16_ycf4* region), the 36 kb inversion (*trnS-GCU_trnG-UCC* region), and the IR region (*ycf1* region) (Fig. 7). As previously pointed out by Magee *et al.* (2010), these fast-evolving regions include gene and intron losses, such as: genes lost from the *rps16_ycf4* region (*rps16, accD, psaI, rpl23 and ycf4*) and introns lost from the *clpP_rps12* and the *rpl16* regions (Fig. 7). This suggests that these regions are most likely involved in structural rearrangements and thus represent unstable regions or hotspots that contribute significantly to the evolutionary dynamics of legume plastomes. Future research on the efficiency of the four classes of nuclear-encoded genes that are involved in chloroplast DNA repair and the maintenance of plastome stability (Maréchal and Brisson, 2010; Guisinger *et al.,* 2011) may reveal if one or several of these four genes are implicated in the legume plastome evolution. Additionally, this study demonstrates that it is essential to sequence plastomes from other papilionoid and earlier legume lineages that remain unexplored to date in order to have a better understanding of the atypical plastome evolution observed in this family.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. Table S1: List of software used in this article. Table S2: Codon usage bias. Table

S3: Repeated elements in *Lupinus luteus* chloroplast genome. Table S4: Sequence divergence (K2P) between *L. luteus* and ten other Fabaceae plastomes. Table S5: Synonymous mutation rate between *L. luteus* and ten other Fabaceae plastome protein-coding sequences. Table S6: Nonsynonymous mutation rate between *L. luteus* and ten other Fabaceae plastomes. Table S7: Ka/Ks ratio between *L. luteus* and ten other Fabaceae plastomes. Table S8: Identification of the presence of inverted repeated elements in *trnS-GGA* and *trnS-GCU* genes within rosid plastomes. Fig. S1: Dot matrix plots showing the presence of a 36 kb inversion in *Lupinus luteus* plastome. Fig. S2: Comparative plastomic maps showing the presence of a 36 kb inversion in *Lupinus luteus* in comparison to other Papilionoideae. Fig. S3: Phylogenetic analysis of plastidic and nuclear *rpl22* protein sequences. Fig. S4: Pairwise distance between *Lupinus luteus* and other Fabaceae orthologous plastomic regions. Fig. S5: Synonymous and nonsynonymous divergence in legume chloroplast *ycf4* gene.

ACKNOWLEDGEMENTS

We thank Prof. J.N. Timmis (University of Adelaide) for his critical reading of the manuscript. This work was supported by UMR-CNRS Ecobio (Rennes, France), and benefited from facilities and support from the "Plate-forme Génomique Environnementale et Fonctionnelle" (OSUR: INEE-CNRS) and the Genouest Bioinformatic Plateform (University of Rennes 1). This work was supported by the Région Bretagne and the European Union Seventh Framework Programme [FP7-CIG-2013-2017; grant number 333709 to M. R-G]. We are greatful to the reviewers for helpful comments and suggestions to improve the early draft of the manuscript.

LITTERATURE CITED

Ainouche A, Bayer RJ. 1999. Phylogenetic relationships in Lupinus (Fabaceae: Papilionoideae) based on internal transcribed spacer sequences (ITS) of nuclear ribosomal DNA. *American journal of botany* **86**: 590-607.

Ainouche A, Bayer RJ, Misset MT. 2004. Molecular phylogeny, diversification and character evolution in Lupinus (Fabaceae) with special attention to Mediterranean and African lupines. *Plant systematics and evolution* **246**: 211-222.

Bartha L, Dragos N, Molnár V, Sramkó G. 2013. Molecular evidence for reticulate speciation in *Astragalus* (Fabaceae) as revealed by a case study from sect. *Dissitiflori*. *Botany* **91**: 702–714.

Bruneau A, Doyle JJ, Palmer JD. 1990. A chloroplast DNA structural mutation as a subtribal character in the Phaseoleae (Leguminosae). *Systematic Botany* **15**: 378-386.

Cai Z, Guisinger M, Kim HG*, et al.* 2008. Extensive reorganization of the plastid genome of Trifolium subterraneum (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *Journal of molecular evolution* **67**: 696-704.

Cardoso D, de Queiroz LP, Pennington RT, e*t al*. 2012. Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. *American journal of botan* **99**: 1991-2013.

Chevreux B, Wetter T, Suhai S. 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information Computer Science and Biology. P*roceedings of the German Conference on Bioinformatics* (GCB) **99**: 45-56.

Clegg M, Zurawski G. 1991. Chloroplast DNA and the study of plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ, eds. *Molecular systematics of plants*. New York: Chapman & Hall.

Clegg MT, Gaut BS, Learn GH Jr, Morton BR. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 6795-6801.

Cosner ME, Raubeson LA, Jansen RK. 2004. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC evolutionary biology* **4**: 27. doi:10.1186/1471-2148-4-27

Crisp M, Gilmore S, van Wyk B. 2000. *Molecular phylogeny of the Genistoid tribes of Papilionoid legumes*. Royal Botanic Gardens, Kew.

Cronk Q, Ojeda I, Pennington RT. 2006. Legume comparative genomics: progress in phylogenetics and phylogenomics. *Current opinion in plant biology* **9**: 99-103.

Cui L, Veeraraghavan N, Richter A, *et al.* 2006. ChloroplastDB: the Chloroplast Genome Database. *Nucleic acids research* **34**: 692-696.

Downie SR, Palmer JD. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ, eds. *Molecular Systematics of Plants*. New-York: Chapman and Hall, 14-35.

Doyle JJ, Doyle JL, Ballenger JA, Palmer JD. 1996. The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. *Molecular phylogenetics and evolution* **5**: 429-438.

Drescher A, Ruf S, Calsa T Jr, Carrer H, Bock R. 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant Journal* **22**: 97-104.

Drummond CS. 2008. Diversification of Lupinus (Leguminosae) in the western New World: derived evolution of perennial life history and colonization of montane habitats. *Molecular phylogenetics and evolution* **48**: 408-421.

Eastwood RJ, Drummond CS, Schifino-Wittmann MT, Hughes CE. 2008. *Diversity and evolutionary history of lupins — Insights from new phylogenies.* Proceedings of the 12th International Lupin Conference — Lupins for health and wealth. Fremantle, Australia.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* **5**: 113. doi:10.1186/1471-2105-5-113

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of molecular biology* **300**: 1005-1016.

Felsenstein J. 1985. Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. *Evolution* **39**: 783-791.

Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. 1991. Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *The EMBO journal* **10**: 3073-3078.

Garcinuno RM, Fernandez-Hernando P, Camara C. 2003. Evaluation of pesticide uptake by Lupinus seeds. *Water research* **37**: 3481-3489.

Guillon F, Champ MM. 2002. Carbohydrate fractions of legumes: uses in human nutrition and potential for health. *The British journal of nutrition* **88**: 293-306.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* **52**: 696-704.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution* **28**: 583-600.

Guo X, Castillo-Ramirez S, Gonzalez V, *et al.* 2007. Rapid evolutionary change of common bean (Phaseolus vulgaris L) plastome, and the genomic diversification of legume chloroplasts. *BMC genomics* **8**: 228. doi:10.1186/1471-2164-8-228

Hughes C, Eastwood R. 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 10334-10339.

Jansen RK, Palmer JD. 1987. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proceedings of the National Academy of Sciences of the United States of America* **84**: 5818-5822.

Jansen RK, Cai Z, Raubeson LA, *et al.* 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 19369-19374.

Jansen RK, Wojciechowski MF, Sanniyasi E, Lee SB, Daniell H. 2008. Complete plastid genome sequence of the chickpea (Cicer arietinum) and the phylogenetic distribution of rps12 and clpP intron losses among legumes (Leguminosae). *Molecular phylogenetics and evolution* **48**: 1204-1217.

Jansen RK, Ruhlman TA. 2012. Plastid genomes of seed plants. In: Bock R, Knoop V, eds. *Genomics of chloroplast and mitochondria.* New York: Springer, 103-126.

Kato T, Kaneko T, Sato S, Nakamura Y, Tabata S. 2000. Complete structure of the chloroplast genome of a legume, Lotus japonicus. *DNA research* **7**: 323-330.

Kazakoff SH, Imelfort M, Edwards D, *et al.* 2012. Capturing the Biofuel Wellhead and Powerhouse: The Chloroplast and Mitochondrial Genomes of the Leguminous Feedstock Tree Pongamia pinnata. *PloS one* **7**: e51687. doi:10.1371/journal.pone.0051687

Kim KJ, Choi KS, Jansen RK. 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Molecular biology and evolution* **22**: 1783-1792.

Kimura M. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**: 275-276.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution* **16**: 111-120.

Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23**: 1026-1028.

Krzywinski M, Schein J, Birol I, *et al.* 2009. Circos: an information aesthetic for comparative genomics. *Genome research* **19**: 1639-1645.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Reasearch* **29**: 4633-4642.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**: R25. doi:10.1186/gb-2009-10-3-r25

Lavin M. Doyle JJ, Palmer JD. 1990. Evolutionary significance of the loss of the chloroplast—DNA inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution* **44**: 390-402.

Lee HL, Jansen RK, Chumley TW, Kim KJ. 2007. Gene relocations within chloroplast genomes of Jasminum and Menodora (Oleaceae) are due to multiple, overlapping inversions. *Molecular biology and evolution* **24**: 1161-1180.

Lewis G, Schrire B, MacKinder B, Lock M. 2005. *Legumes of the world*. *Royal Botanical Gardens*, Kew.

Magee AM, Aspinall S, Rice DW, *et al.* 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research* **20**: 1700-1710.

Magni C, Sessa F, Accardo E, *et al.* 2004. Conglutin gamma, a lupin seed protein, binds insulin in vitro and reduces plasma glucose levels of hyperglycemic rats. *The Journal of nutritional biochemistry* **15**: 646-650.

Mahé F, Markova D, Pasquet R, Misset MT, Ainouche A. 2011a. Isolation, phylogeny and evolution of the SymRK gene in the legume genus *Lupinus* L. *Molecular phylogenetics and evolution* **60**: 49-61.

Mahé F, Pascual H, Coriton O, *et al.* 2011b. New data and phylogenetic placement of the enigmatic old world lupin: Lupinus mariae-josephi H. Pascual. *Genetic resources and crop evolution* **58**: 101-114.

Maul JE, Lilly JW, Cui L, *et al.* 2002. The Chlamydomonas reinhardtii plastid chromosome: islands of genes in a sea of repeats. *Plant Cell* **14**: 2659-2679.

Maréchal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. *New Phytologist* **186**: 299-317.

Millen RS, Olmstead RG, Adams KL, *et al.* 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* **13**: 645-658.

Moore MJ, Bell CD, Soltis PS, Soltis DE. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 19363-19368.

Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 4623-4628.

Palmer JD, Thompson WF. 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* **29**: 537-50.

Palmer JD. 1985. Comparative organization of chloroplast genomes. *Annual Review of Genetics* **19**: 325–354.

Palmer JD, Herbon LA. 1988. Plant mitochondrial-DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution* **28**: 87-97.

Paradis E, Bolker B, Claude J, *et al.* 2011. available at: http://cran.r-project.org/web/packages/ape/ape.pdf. Package 'ape'.

Perry AS, Brennan S, Murphy DJ, Kavanagh TA, Wolfe KH. 2002. Evolutionary re-organisation of a large operon in adzuki bean chloroplast DNA caused by inverted repeat movement. *DNA research* **9**: 157-162.

Pierleoni A, Martelli PL, Fariselli P, Casadio R. 2006. BaCelLo: a balanced subcellular localization predictor. *Bioinformatics* **22**: e408-16.

Pilvi TK, Jauhiainen T, Cheng ZJ, Mervaala EM, Vapaatalo H, Korpela R. 2006. Lupin protein attenuates the development of hypertension and normalises the vascular function of NaCl-loaded Goto-Kakizaki rats. *Journal of physiology and pharmacology : an official journal of the Polish Physiological Society* **57**: 167-176.

Pombert JF, Otis C, Lemieux C, Turmel M. 2005. The chloroplast genome sequence of the green alga Pseudendoclonium akinetum (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. *Molecular Biology and Evolution* **22**: 1903-1918.

Raubeson LA, Jansen RK. 2005. *Chloroplast genomes of plants*. Oxfordshire: CABI Publishing.

Raubeson LA, Peery R, Chumley TW, *et al.* 2007. Comparative chloroplast genomics: analyses including new sequences from the angiosperms Nuphar advena and Ranunculus macranthus. *BMC genomics* **8**: 174. doi:10.1186/1471-2164-8-174

Rogalski M, Ruf S, Bock R. 2006. Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic acids research* **34**: 4537-4545.

Ruf S, Kössel H, Bock R. 1997. Targeted inactivation of a tobacco intron-containing open reading frame reveals a novel chloroplast-encoded photosystem I-related gene. *Journal of cell biology* **139**: 95-102.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**: 406-425.

Saski C, Lee SB, Daniell H, *et al.* 2005. Complete chloroplast genome sequence of Glycine max and comparative analyses with other legume genomes. *Plant molecular biology* **59**: 309-322.

Small I, Peeters N, Legeai F, Lurin C. 2004. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**: 1581-1590.

Sprent JI. 2007. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytologist* **174**: 11-25.

Stefanovic S, Pfeil BE, Palmer JD, Doyle JJ. 2009. Relationships among phaseolid legumes based on sequences from eight chloroplast regions. *Systematic Botany* **34**: 115–128.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**: 2731-2739.

Tangphatsornruang S, Sangsrakru D, Chanprasert J*, et al.* 2010. The chloroplast genome sequence of mungbean (Vigna radiata) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. *DNA research* **17**: 11-22.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Review Genetics* **5**: 123-135.

Tsumura Y, Suyama Y, Yoshimura K. 2000. Chloroplast DNA inversion polymorphism in populations of Abies and Tsuga. *Molecular biology and evolution* **17**: 1302-1312.

Ueda M, Fujimoto M, Arimura SI, Tsutsumi N, Kadowaki KI. 2008. Presence of a latent mitochondrial targeting signal in gene on mitochondrial genome. *Molecular Biology and Evolution* **25**: 1791-1793.

Wojciechowski MF, Lavin M, Sanderson MJ. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported subclades within the family. *American Journal of Botany* **91**: 1846-1862.

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**: 3252-3255.

Yang Z, Nielsen R. 2000. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Molecular Biology and Evolution* **17,** 32–43.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**: 1586-91.

Figure

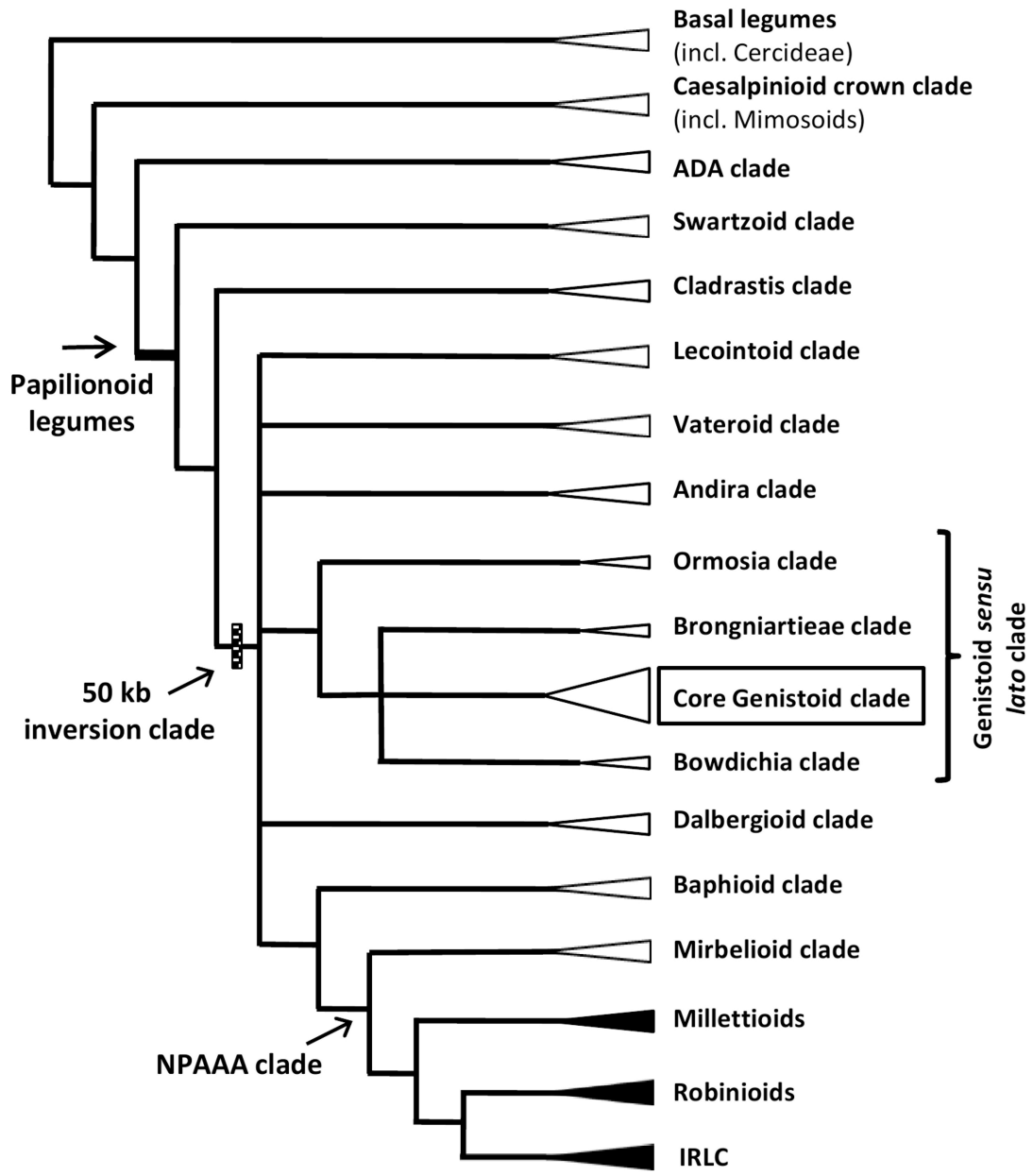

Fig. 1 Simplified phylogenetic tree representing the main clades circumscribed in the legume family (redrawn from Wojciechowski *et al.*, 2004 and Cardoso *et al.*, 2012). Black triangles indicate lineages for which one or more whole chloroplast genome sequences are available, whereas open triangles indicate lineages for which no plastome sequence exists, including the core Genistoid clade (boxed) targeted in this study.

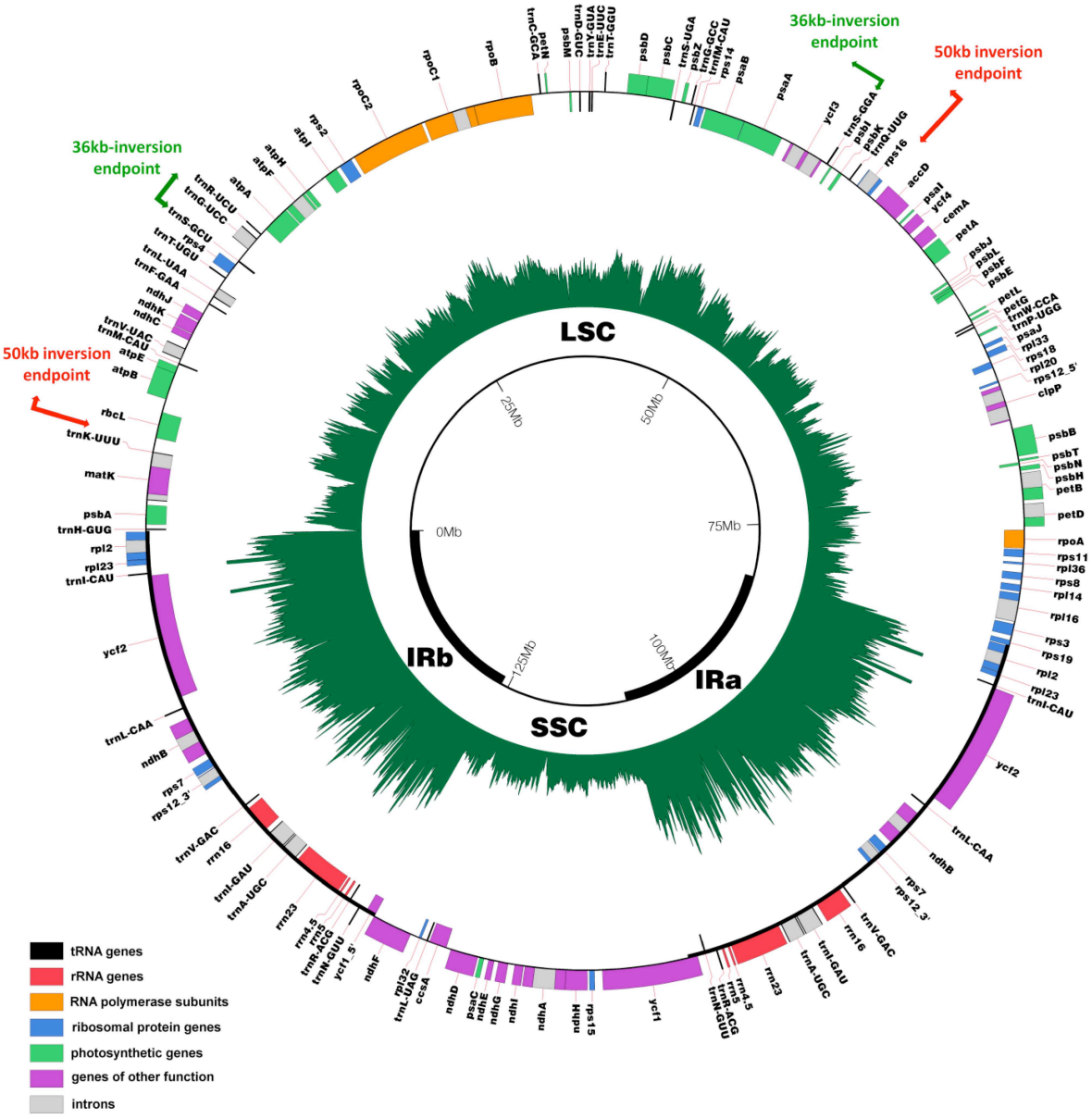

Fig. 2 Circular gene map of *Lupinus luteus* (Genistoid; Fabaceae) plastid genome. Genes are represented with boxes inside and outside the first circle to indicate clockwise or counterclockwise transcription direction, respectively. Genes belonging to different functional groups are color coded. Read depth coverage of the plastome is represented by proportional radial lines in the inner green circle. The location of the different main plastomic regions (Inverted Repeats, Large Single Copy and Small Single Copy) are indicated in the inner circle. The endpoints of the 50 kb and 36 kb inversions are represented by red and green arrows, respectively.

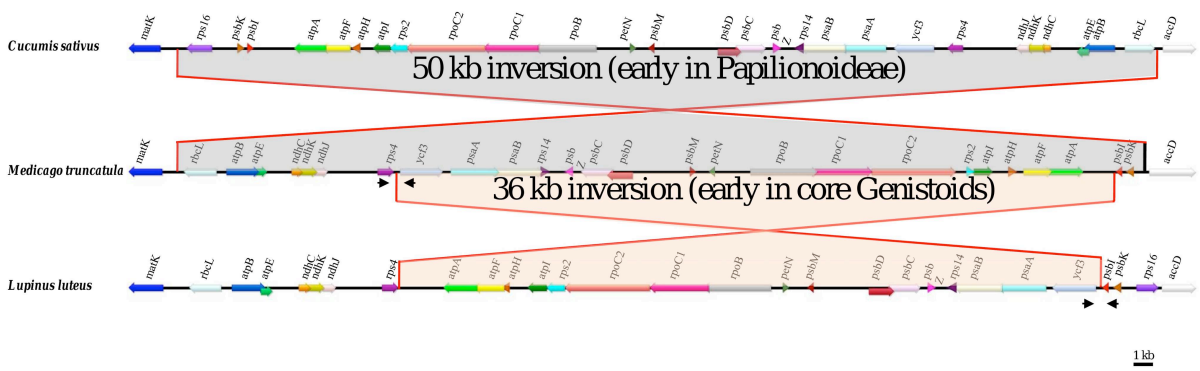

Fig. 3 Comparative plastomic maps showing the endpoints of the large 50 kb inversion present in most Papilionideae (Fabaceae) and of a new 36 kb inversion detected in most core Genistoids surveyed in this study. The plastomes of *Cucumis sativus*, *Medicago truncatula* and *Lupinus luteus* were used to represent the structural patterns observed in most flowering plants, in most Papilionoid legumes and in the core Genistoids, respectively. The partial plastomic maps are drawn to scale, and only protein-coding genes are mapped. Approximate position of diagnostic primer pairs used to detect the presence or absence of the 36 kb inversion are designated by black arrows (not to scale).

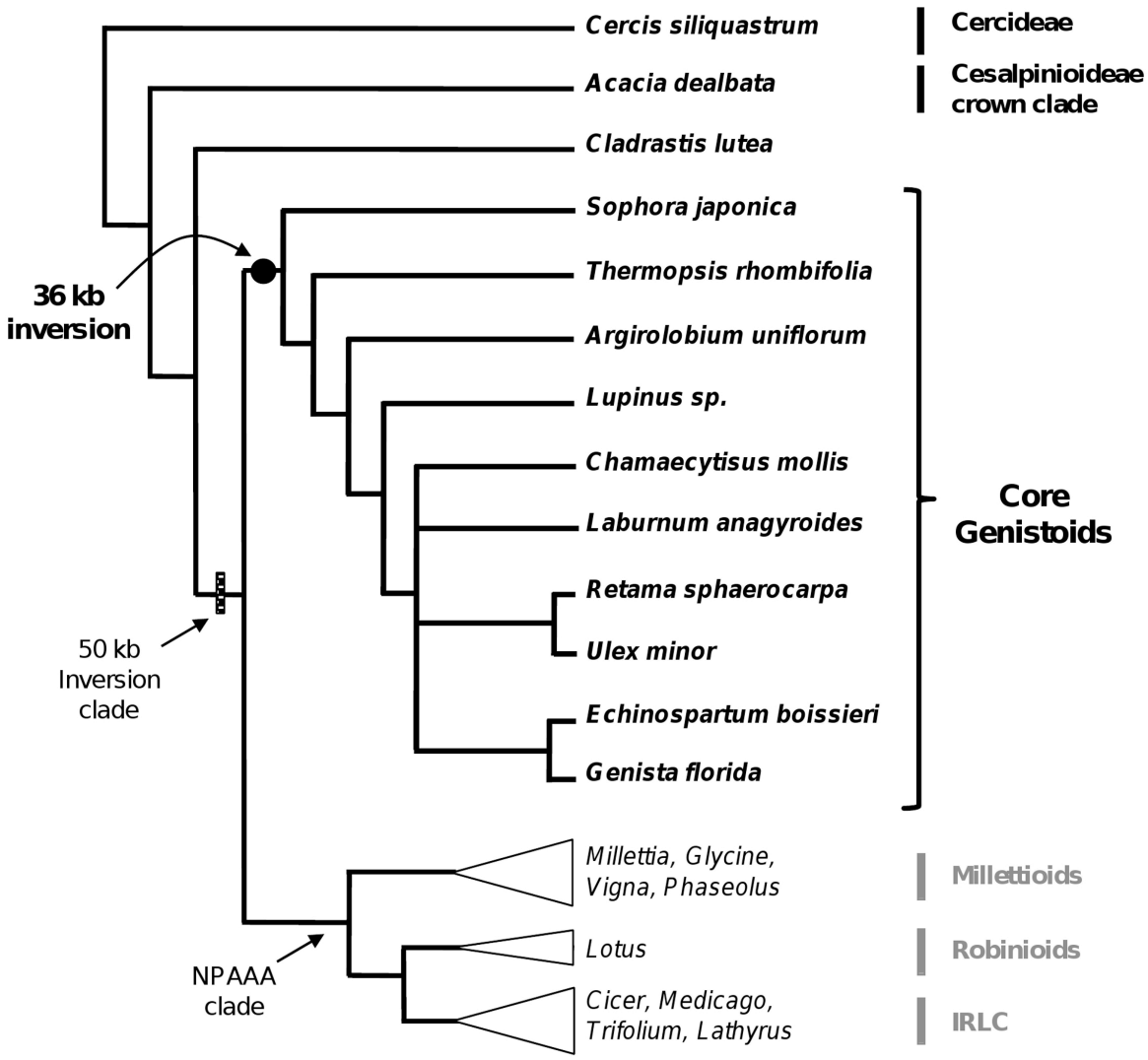

Fig. 4 Phylogenetic position of the 36 kb inversion rearrangement (solid black circle) detected in the plastomes of *Lupinus* and representatives of the core Genistoid clade (Papilionoidae; Fabaceae). All taxa screened by PCR and sequencing for the presence or absence of this inversion have their names labelled in bold. The taxa for which the plastome sequence is publicly available and for which the 36kb inversion is absent belong to the Millettioids, Robinioids and IRLC (labelled in gray). The phylogenetic tree is redrawn from Cardoso *et al.* (2012).

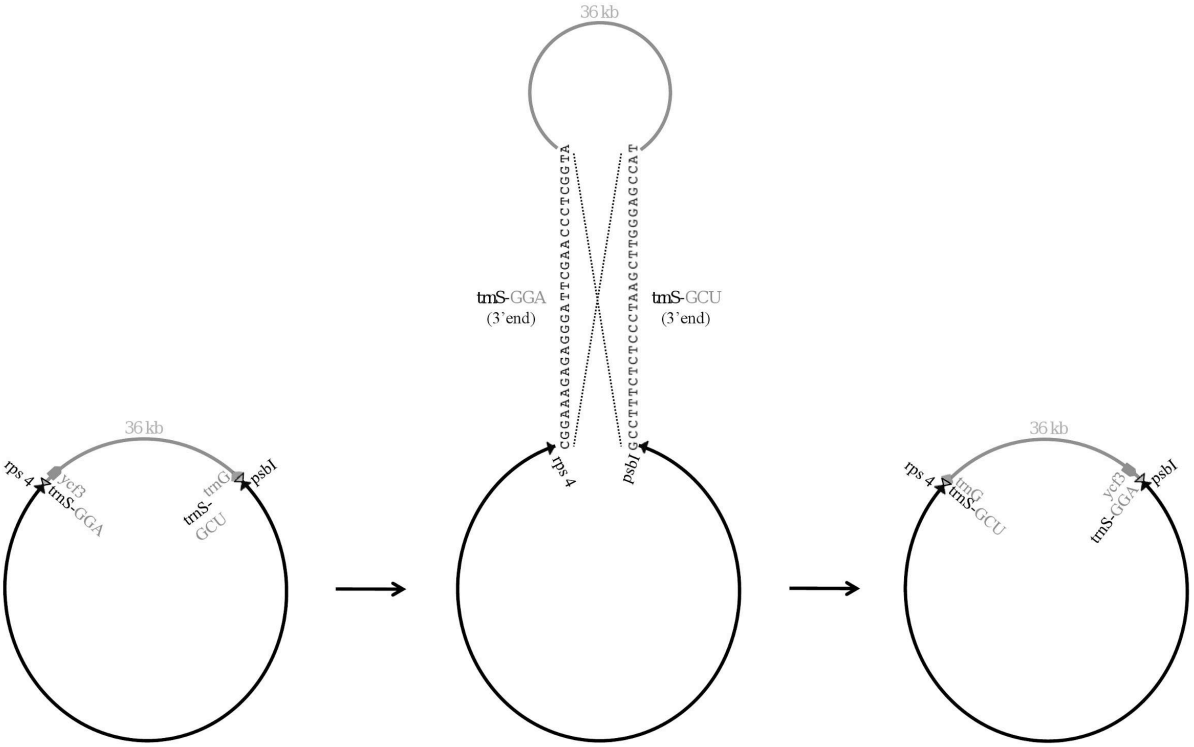

Fig. 5 Illustration of the suggested flip-flop recombination event that occurred after the divergence of the Genistoids from the other Fabaceae clades, resulting in a 36 kb inversion (gray region). This recombination event was most likely caused by the presence of inverted repeat sequences (29 bp shown) at the 3'end of the *trnS-GGA* and *trnS-GCU* that are 36 kb apart.

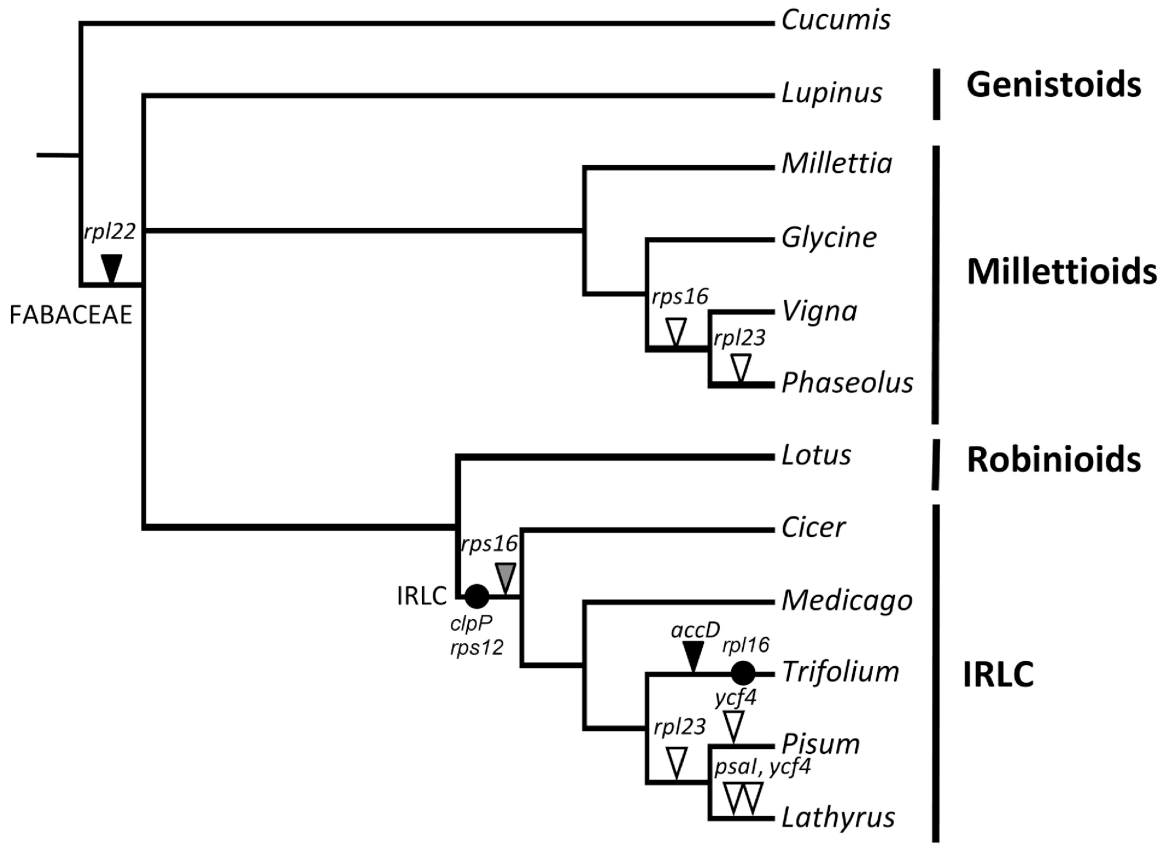

Fig. 6 Chloroplast genes and introns lost during Fabaceae evolution. All Fabaceae species whose plastome has been fully sequenced are presented in the phylogenetic tree (redrawn from Cardoso *et al.*, 2012). Black circles on branches indicate intron losses whereas triangles show the genes recently lost from the plastome in various Fabaceae species. Black triangles indicate genes that were shown to be functionally transferred to the nucleus (Gantt *et al.,* 1991; Magee *et al.*, 2010), including the transfer of *rpl*22 in *Lupinus luteus* detected in this study. The gray triangle indicates the functional replacement of the plastomic *rps*16 gene by a mitochondrial gene functionally transferred to the nucleus (Ueda *et al.*, 2008). White triangles show the plastid genes lost during Fabaceae evolution and for which no functional replacement in the nucleus has been observed. The independent losses of *ycf4* in *Lathyrus sativus* and *Pisum sativum* were inferred from Magee *et al.* (2010) results who showed the presence of intact *ycf4* genes in some *Lathyrus* species.

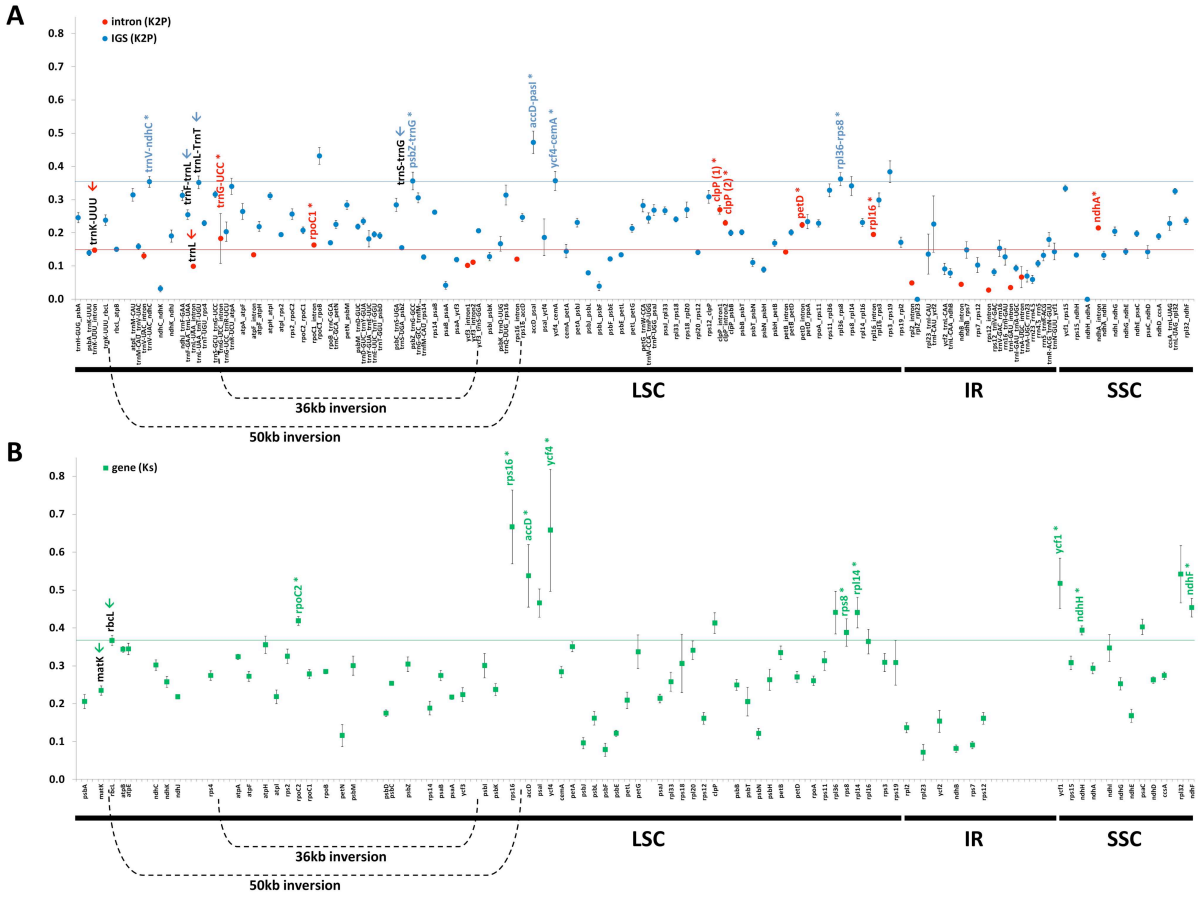

Fig. 7 Mean sequence divergence ± standard error between homologous regions of the *Lupinus luteus* and other legume plastomes. The *x*- axis lists intronic, intergenic and protein-coding regions in the same order than in *L. luteus* plastome. (A) Red and blue solid circles show mean sequence divergence for each orthologous intronic or intergenic pair, calculated using K2p model (Kimura, 1980). (B) Green solid square show mean sequence divergence for each orthologous protein coding gene pair, estimated with the synonymous mutation rate (Ks) and using yn00 method (Yang, 2007). The various regions (intronic, intergenic or protein-coding gene pair) previously used in Fabaceae phylogenetic studies are indicated with red, blue and green arrows, respectively. The intronic, intergenic or protein-coding regions presenting a higher evolutionary rate than those previously used in Fabaceae evolutionary studies and presenting a minimum size of 300 bp are indicated with red, blue or green stars, respectively.

Table 1 *Lupinus luteus* plastome characteristics

Plastome characteristics	***Lupinus luteus***
Size (bp)	151,894
LSC size in bp (%)	82,327 (54.2)
SSC size in bp (%)	17,847 (11.7)
IR length in bp (%)	25,860 (34.1)
Size in bp (%) coding regions	90,217 (59.4)
Size in bp (%) of protein-coding regions	78,363 (51.6)
Size in bp (%) of introns	19,136 (12.6)
Size in bp (%) of rRNA	9,056 (6)
Size in bp (%) of tRNA	2,798 (1.8)
Size in bp (%) of IGS	42,541 (28)
Number of different genes	111
Number of different protein-coding genes	77
Number of different tRNA genes	30
Number of different rRNA genes	4
Number of different genes duplicated by IR	17
Number of different genes with introns	18
Overall % GC content[a]	36.6
% GC content in protein-coding regions[a]	37.3
% GC content in introns[a]	36.3
% GC content in IGS[a]	30.3
% GC content in rRNA[a]	55.3
% GC content in tRNA[a]	53.3

[a] The sequence of the two Inverted Repeats were taken into account for this analysis.

Table S1: List of software used in this paper. The purpose and availability of each software are mentionned.

Software	Purpose	Availability
BaCelLo (Pierleoni *et al.*, 2006)	predict subcellular localization of proteins	http://gpcr.biocomp.unibo.it/bacello/pred.htm
Bowtie (Langmead *et al.*, 2009)	reference-guided assembly/ read mapping	https://github.com/BenLangmead/bowtie
CIRCOS (Krzywinski *et al.*, 2009)	visualization tool for comparative genomics	http://circos.ca/
DOGMA (Wyman *et al.*, 2004)	automatic anntotation of organellar genomes	http://dogma.ccbb.utexas.edu/
Gepard (Krumsiek *et al.*, 2007)	create dotplots on genome scale	http://www.helmholtz-muenchen.de/icb/gepard
MIRA (Chevreux *et al.*, 1999)	de novo assembly, mapping	http://www.chevreux.org/projects_mira.html
MUSCLE (Edgar, 2004)	multiple sequence alignment	http://www.drive5. com/muscle
Newbler (454 Life Science)	de novo DNA sequence assembly	http://www.454.com/products/analysis-software/
PAML (Yang, 2007)	phylogenetic analysis by maximum likelihood	http://abacus.gene.ucl.ac.uk/software/paml.html
Predotar (Small *et al.*, 2004)	predict subcellular localization of proteins	http://urgi.versailles.inra.fr/predotar/predotar.html
REPuter (Kurtz *et al.*, 2001)	identification of repeat elements	http://bibiserv.techfak.uni-bielefeld.de/reputer/
TargetP (Emanuelsson *et al.*, 2000)	predict subcellular localization of proteins	http://www.cbs.dtu.dk/services/TargetP/

Table S2. Codon usage bias. For each amino acid where synonymous codons exists**,** frequencies of codons ending by A or T and C or G bases were compared to the expected frequencies in case of no bias of codon usage. A Chi² test corrected for multiple testing using a bonferroni correction was performed.

	Third position nucleotide	Codon number	Expected frequency	P-value (bonferroni corrected)	X²
Phe	A-T	923	0,50	<0,001	X-squared = 204.8611, df = 1, p-value < 2.2e-16
	G-C	402	0,50		
Leu	A-T	1601	0,50	<0,001	X-squared = 269.5646, df = 1, p-value < 2.2e-16
	G-C	797	0,50		
Ile	A-T	1655	0,67	<0,001	X-squared = 204.7166, df = 1, p-value < 2.2e-16
	G-C	372	0,33		
Val	A-T	935	0,50	<0,001	X-squared = 349.0181, df = 1, p-value < 2.2e-16
	G-C	283	0,50		
Ser	A-T	1162	0,50	<0,001	X-squared = 251.4633, df = 1, p-value < 2.2e-16
	G-C	513	0,50		
Pro	A-T	593	0,50	<0,001	X-squared = 78.9106, df = 1, p-value < 2.2e-16
	G-C	324	0,50		
Thr	A-T	796	0,50	<0,001	X-squared = 184.9771, df = 1, p-value < 2.2e-16
	G-C	338	0,50		
Ala	A-T	913	0,50	<0,001	X-squared = 256.4579, df = 1, p-value < 2.2e-16
	G-C	345	0,50		
Tyr	A-T	685	0,50	<0,001	X-squared = 358.276, df = 1, p-value < 2.2e-16
	G-C	141	0,50		
His	A-T	393	0,50	<0,001	X-squared = 141.2791, df = 1, p-value < 2.2e-16
	G-C	123	0,50		
Gln	A-T	633	0,50	<0,001	X-squared = 286.8051, df = 1, p-value < 2.2e-16
	G-C	157	0,50		
Asn	A-T	850	0,50	<0,001	X-squared = 366.383, df = 1, p-value < 2.2e-16
	G-C	223	0,50		
Lys	A-T	963	0,50	<0,001	X-squared = 362.8403, df = 1, p-value < 2.2e-16
	G-C	289	0,50		
Asp	A-T	704	0,50	<0,001	X-squared = 316.8, df = 1, p-value < 2.2e-16
	G-C	176	0,50		
Glu	A-T	878	0,50	<0,001	X-squared = 306.2207, df = 1, p-value < 2.2e-16
	G-C	282	0,50		
Cys	A-T	209	0,50	<0,001	X-squared = 81.1259, df = 1, p-value < 2.2e-16
	G-C	61	0,50		
Arg	A-T	974	0,50	<0,001	X-squared = 321.7594, df = 1, p-value < 2.2e-16
	G-C	327	0,50		
Gly	A-T	1141	0,50	<0,001	X-squared = 368.3873, df = 1, p-value < 2.2e-16
	G-C	390	0,50		
STOP	A-T	62	0,67	NS	X-squared = 6.6494, df = 1, p-value = 0.009919
	G-C	15	0,33		

Table S3: Repeated elements in the chloroplast genome of *Lupinus luteus*.

Position of the first repeat		Position of the second repeat					
Start	Localisation	Start	Localisation	Nb. of mismatches between repeats	Repetition type	Motif size	Distance between repeated elements (bp)
283	Intergenic (trnH_GUG-psbA)	307	Intergenic (trnH_GUG-psbA)	0	Palindrome	24	-
15939	Genic (trnS-GCU)	52078	Genic (trnS-GGA)	2	Inverted	35	36104
16116	Intergenic (trnS_GCU-trnG_UCC)	16163	Intergenic (trnS_GCU-trnG_UCC)	0	Palindrome	47	-
42874	Genic (trnS-UGA)	52083	Genic (trnS-GGA)	3	Inverted	32	9177
43663	Intergenic (psbZ-trnG_GCC)	43688	Intergenic (psbZ-trnG_GCC)	1	Palindrome	25	-
46074	Genic (psaB)	48298	Genic (psaA)	5	Direct	52	2224
50919	Intron (ycf3)	79871	Intron (rpl16)	3	Direct	39	28952
50919	Intron (ycf3)	116634	Intron (ndhA)	3	Inverted	39	65676
64194	Intergenic (trnW_CCA-trnP_UGG)	64215	Intergenic (trnW_CCA-trnP_UGG)	1	Palindrome	17	-
65063	Intergenic (psaJ-rpl33)	65101	Intergenic (psaJ-rpl33)	2	Palindrome	30	-
65334	Intergenic (psaJ-rpl33)	109295	Genic (ycf1)	3	Inverted	30	43931
66338	Intergenic (rps18-rpl20)	66371	Intergenic (rps18-rpl20)	1	Palindrome	33	-
72118	Intergenic (psbT-psbN)	72143	Intergenic (psbT-psbN)	1	Palindrome	23	-
73538	Intron (petB)	79876	Intron (rpl16)	3	Inverted	31	6307
73539	Intron (petB)	116637	Intron (ndhA)	3	Direct	31	43098
79859	Intron (rpl16)	116634	Intron (ndhA)	1	Inverted	51	36724
79871	Intron (rpl16)	96868	Intergenic (rps12-trnV_GAC)	1	Non inverted	40	16997
84309	Intergenic (rpl23-trnI_CAU)	86294	Genic (ycf2)	0	Non inverted	288	1985
87308	Genic (ycf2)	87344	Genic (ycf2)	2	Tandem	40	-
87308	Genic (ycf2)	146839	Genic (ycf2)	2	Inverted	40	59491
87344	Genic (ycf2)	146875	Genic (ycf2)	2	Inverted	40	59491
88531	Genic (ycf2)	88558	Genic (ycf2)	2	Tandem	29	-
88531	Genic (ycf2)	145634	Genic (ycf2)	2	Inverted	31	57072
88558	Genic (ycf2)	145661	Genic (ycf2)	2	Inverted	31	57072
89887	Genic (ycf2)	89905	Genic (ycf2)	1	Tandem	18	-
89894	Genic (ycf2)	144285	Genic (ycf2)	3	Inverted	35	54356
89908	Genic (ycf2)	144298	Genic (ycf2)	0	Inverted	35	54355
96866	Intergenic (rps12-trnV_GAC)	116634	Intron (ndhA)	1	Inverted	41	19727
110395	Genic (ycf1)	110416	Genic (ycf1)	1	Palindrome	16	-
116634	Intron (ndhA)	137316	Intergenic (trnV_GAC-rps12)	1	Direct	41	20682
122951	Intergenic (trnL_UAG-rpl32)	122977	Intergenic (trnL_UAG-rpl32)	1	Palindrome	25	-

Table S4. Sequence divergence (K2P) between *L. luteus* and ten other Fabaceae plastomes. For gene with exons, a unique value has been calculated for the whole coding region.

region	type	size (pb)	*Phaseolus vulgaris*	*Pisum sativum*	*Vigna radiata*	*Glycine max*	*Lathyrus sativus*	*Cicer arietinum*	*Trifolium subterraneum*	*Medicago truncatula*	*Lotus japonicus*	*Milletia pinnata*	*mean*	*standard deviation*	*standard error*
trnH-GUG	tRNA	75	0,000	0,000	0,000	0,000	0,000	0,014	0,014	0,014	0,000	0,000	0,004	0,007	0,002
trnH-GUG_psbA	IGS	240	0,175	0,264	0,163	0,211	0,249	0,306	0,252	0,237	0,330	0,274	0,246	0,053	0,017
psbA	CDS	1062	0,056	0,031	0,055	0,048	0,031	0,033	0,044	0,035	0,039	0,047	0,042	0,010	0,003
psbA_trnK-UUU	IGS	275	0,155	0,142	0,182	0,123	0,175	0,123		0,142	0,110	0,110	0,140	0,026	0,009
trnK-UUU	tRNA	35	0,044	0,014	0,044	0,014	0,014	0,014	0,014	0,014	0,014	0,029	0,021	0,012	0,004
trnK-UUU_intron	INTRON	318	0,193	0,139	0,191	0,161	0,137	0,123	0,131	0,123	0,147	0,137	0,148	0,026	0,008
matK	CDS	1521	0,200	0,140	0,202	0,169	0,133	0,124	0,128	0,121	0,144	0,142	0,150	0,030	0,009
trnK-UUU_intron	INTRON	773	0,193	0,139	0,191	0,161	0,137	0,123	0,131	0,123	0,147	0,137	0,148	0,026	0,008
trnK-UUU	tRNA	37	0,044	0,014	0,044	0,014	0,014	0,014	0,014	0,014	0,014	0,029	0,021	0,012	0,004
trnK-UUU_rbcL	IGS	787	0,297	0,208	0,303	0,261	0,201	0,149	0,182	0,236	0,259	0,287	0,238	0,052	0,017
rbcL	CDS	1428	0,061	0,073	0,072	0,069	0,072	0,066	0,081	0,079	0,068	0,062	0,070	0,007	0,002
rbcL_atpB	IGS	767	0,186	0,158	0,188	0,126	0,158	0,148	0,141	0,131	0,148	0,126	0,151	0,022	0,007
atpB	CDS	1495	0,089	0,092	0,088	0,075	0,097	0,085	0,100	0,091	0,076	0,075	0,087	0,009	0,003
atpE	CDS	400	0,100	0,111	0,097	0,086	0,117	0,088	0,094	0,093	0,105	0,094	0,098	0,010	0,003
atpE_trnM-CAU	IGS	139	0,295	0,279	0,386	0,301	0,277	0,237	0,261	0,299	0,449	0,362	0,315	0,065	0,021
trnM-CAU	tRNA	74	0,000	0,017	0,000	0,000	0,017	0,017	0,017	0,017	0,017	0,000	0,010	0,009	0,003
trnM-CAU_trnV-UAC	IGS	202	0,175	0,165	0,205	0,148	0,185	0,103	0,157	0,121	0,157	0,177	0,159	0,030	0,009
trnV-UAC	tRNA	39	0,053	0,053	0,053	0,000	0,053	0,053	0,053	0,053	0,071	0,053	0,049	0,018	0,006
trnV-UAC_intron	INTRON	580	0,173	0,114	0,193	0,130	0,119	0,116	0,108	0,110	0,094	0,153	0,131	0,032	0,010
trnV-UAC	tRNA	36	0,053	0,053	0,053	0,000	0,053	0,053	0,053	0,053	0,071	0,053	0,049	0,018	0,006
trnV-UAC_ndhC	IGS	497	0,441	0,276	0,396	0,382	0,301	0,399		0,327	0,343	0,327	0,355	0,053	0,018
ndhC	CDS	363	0,075	0,088	0,081	0,066	0,085	0,082	0,088	0,091	0,082	0,070	0,081	0,008	0,003
ndhC_ndhK	IGS	43	0,058	0,000	0,058	0,058	0,000	0,000		0,000	0,059	0,058	0,032	0,031	0,010
ndhK	CDS	690	0,088	0,080	0,088	0,086	0,083	0,096		0,075	0,082	0,073	0,083	0,007	0,002
ndhK_ndhJ	IGS	108	0,271	0,097	0,249	0,272	0,132	0,169	0,189	0,132	0,189	0,209	0,191	0,060	0,019
ndhJ	CDS	477	0,073	0,073	0,073	0,059	0,066	0,064	0,057	0,054	0,052	0,059	0,063	0,008	0,003
ndhJ_trnF-GAA	IGS	588	0,332	0,259	0,375	0,331	0,314	0,290	0,277	0,261	0,416	0,278	0,313	0,051	0,016
trnF-GAA	tRNA	73	0,014	0,014	0,029	0,014	0,043	0,029	0,014	0,014	0,014	0,000	0,018	0,012	0,004
trnF-GAA_trnL-UAA	IGS	406	0,338	0,214	0,273	0,234	0,273	0,178	0,255	0,217	0,273	0,295	0,255	0,046	0,015
trnL-UAA	tRNA	48	0,208	0,189	0,226	0,043	0,189	0,207	0,189	0,189	0,189	0,189	0,182	0,050	0,016
trnL-UAA_intron	INTRON	501	0,122	0,086	0,122	0,087	0,102	0,087	0,098	0,094	0,094	0,102	0,100	0,013	0,004
trnL-UAA	tRNA	31	0,208	0,189	0,226	0,043	0,189	0,207	0,189	0,189	0,189	0,189	0,182	0,050	0,016

trnL-UAA_trnT-UGU	IGS	624	0,456	0,320	0,444	0,386	0,303	0,341		0,315	0,298	0,307	0,352	0,062	0,021
trnT-UGU	tRNA	72	0,031	0,000	0,031	0,015	0,015	0,000	0,047	0,000	0,000	0,000	0,014	0,017	0,005
trnT-UGU_rps4	IGS	269	0,244	0,233	0,206	0,188	0,234	0,253	0,235	0,284	0,224	0,196	0,230	0,028	0,009
rps4	CDS	612	0,115	0,105	0,115	0,094	0,121	0,096	0,114	0,082	0,102	0,114	0,106	0,012	0,004
rps4_trnS-GCU	IGS	207													
trnS-GCU	tRNA	87	0,052	0,026	0,052	0,026	0,026	0,013	0,013	0,013	0,013		0,026	0,016	0,005
trnS-GCU_trnG-UCC	IGS	782	0,357	0,314		0,297	0,310				0,303		0,316	0,024	0,011
trnG-UCC	tRNA	23	0,302	0,016	0,302	0,000	0,016				0,000	0,000	0,091	0,144	0,055
trnG-UCC_intron	INTRON	698	0,587	0,119		0,088	0,108				0,122	0,077	0,183	0,199	0,081
trnG-UCC	tRNA	49	0,302	0,016	0,302	0,000	0,016				0,000	0,000	0,091	0,144	0,055
trnG-UCC_trnR-UCU	IGS	294	0,242	0,219		0,152	0,244				0,292	0,072	0,204	0,079	0,032
trnR-UCU	tRNA	72	0,014	0,000	0,014	0,014	0,014	0,044	0,000	0,000	0,029	0,014	0,014	0,014	0,004
trnR-UCU_atpA	IGS	353	0,345	0,488	0,348	0,211	0,443	0,363	0,363	0,291	0,276	0,275	0,340	0,082	0,026
atpA	CDS	1533	0,069	0,076	0,070	0,059	0,086	0,079	0,080	0,077	0,070	0,067	0,073	0,008	0,002
atpA_atpF	IGS	65	0,395	0,214	0,350	0,289	0,245	0,154	0,245	0,184	0,353	0,218	0,265	0,079	0,025
atpF	CDS	411	0,127	0,122	0,124	0,095	0,122	0,115	0,094	0,106	0,104	0,106	0,112	0,012	0,004
atpF_intron	INTRON	742	0,160	0,127	0,168	0,113	0,145	0,131	0,121	0,123	0,123	0,131	0,134	0,018	0,006
atpF	CDS	144	0,127	0,122	0,124	0,095	0,122	0,115	0,094	0,106	0,104	0,106	0,112	0,012	0,004
atpF_atpH	IGS	70	0,201	0,179	0,201	0,179	0,225	0,298	0,179	0,178	0,249	0,303	0,219	0,049	0,015
atpH	CDS	246	0,073	0,096	0,068	0,073	0,100	0,105	0,091	0,096	0,064	0,051	0,082	0,018	0,006
atpH_atpI	IGS	1139	0,352	0,322	0,379	0,284	0,330	0,287	0,283	0,307	0,303	0,272	0,312	0,034	0,011
atpI	CDS	744	0,055	0,062	0,053	0,063	0,062	0,068	0,065	0,067	0,076	0,055	0,063	0,007	0,002
atpI_rps2	IGS	255	0,202	0,192	0,202	0,192	0,201	0,201	0,201	0,173	0,221	0,164	0,195	0,016	0,005
rps2	CDS	711	0,089	0,060	0,089	0,061	0,126	0,046	0,096	0,033	0,054	0,060	0,071	0,028	0,009
rps2_rpoC2	IGS	229	0,326	0,271	0,339	0,240	0,248	0,221	0,259	0,219	0,283	0,163	0,257	0,052	0,016
rpoC2	CDS	4149	0,146	0,136	0,147	0,131	0,144	0,169	0,137	0,125	0,143	0,127	0,141	0,013	0,004
rpoC2_rpoC1	IGS	206	0,217	0,197	0,198	0,178	0,217	0,217		0,228	0,270	0,150	0,208	0,034	0,011
rpoC1	CDS	1614	0,087	0,094	0,092	0,071	0,113	0,085	0,102	0,089	0,091	0,071	0,090	0,013	0,004
rpoC1_intron	INTRON	766	0,186	0,161	0,198	0,151	0,154	0,168	0,158	0,162	0,159	0,140	0,164	0,017	0,005
rpoC1	CDS	432	0,087	0,094	0,092	0,071	0,113	0,085	0,102	0,089	0,091	0,071	0,090	0,013	0,004
rpoC1_rpoB	IGS	26	0,398	0,548	0,334	0,339	0,398	0,406	0,548	0,406	0,548	0,398	0,432	0,084	0,027
rpoB	CDS	3213	0,096	0,087	0,089	0,082	0,102	0,084	0,099	0,082	0,092	0,080	0,089	0,007	0,002
rpoB_trnC-GCA	IGS	476	0,184	0,162	0,184	0,154	0,170	0,179		0,193	0,141	0,168	0,171	0,016	0,005
trnC-GCA	tRNA	80	0,062	0,062	0,062	0,062	0,078	0,062	0,078	0,096	0,062	0,062	0,068	0,012	0,004
trnC-GCA_petN	IGS	277	0,276	0,262	0,241	0,264		0,170	0,182	0,194	0,218	0,226	0,226	0,038	0,013
petN	CDS	90	0,082	0,046	0,070	0,070	0,047	0,034	0,083	0,046	0,047	0,046	0,057	0,017	0,006
petN_psbM	IGS	1189	0,295		0,328	0,275	0,338	0,257	0,313	0,280	0,261	0,213	0,284	0,039	0,013
psbM	CDS	105	0,060	0,049	0,060	0,072	0,060	0,060	0,049	0,081	0,081	0,050	0,062	0,012	0,004
psbM_trnD-GUC	IGS	437	0,248	0,215	0,268	0,207	0,240	0,190	0,223	0,199	0,208	0,193	0,219	0,026	0,008
trnD-GUC	tRNA	74	0,014	0,014	0,014	0,000	0,014	0,014	0,014	0,014	0,014	0,000	0,011	0,006	0,002
trnD-GUC_trnY-GUA	IGS	458	0,269	0,239	0,276	0,243	0,240	0,162	0,225	0,215	0,262	0,222	0,235	0,033	0,010
trnY-GUA	tRNA	84	0,000	0,000	0,000	0,000	0,000	0,012	0,012	0,000	0,000	0,000	0,002	0,005	0,002
trnY-GUA_trnE-UUC	IGS	60	0,370	0,150	0,262	0,218	0,171	0,171	0,150	0,110	0,110	0,110	0,182	0,082	0,026

trnE-UUC	tRNA	73	0,014	0,043	0,000	0,000	0,043	0,028	0,028	0,028	0,014	0,000	0,020	0,017	0,005
trnE-UUC_trnT-GGU	IGS	635	0,218	0,163	0,227	0,211	0,210	0,156	0,203	0,179	0,227	0,164	0,196	0,028	0,009
trnT-GGU	tRNA	72	0,048	0,066	0,048	0,032	0,032	0,048	0,048	0,032	0,032	0,048	0,043	0,011	0,004
trnT-GGU_psbD	IGS	1191	0,247	0,162	0,242	0,197	0,183	0,166	0,184	0,162	0,189	0,185	0,192	0,030	0,010
psbD	CDS	1033	0,052	0,048	0,046	0,035	0,043	0,043	0,048	0,042	0,040	0,042	0,044	0,005	0,002
psbC	CDS	1398	0,053	0,057	0,059	0,051	0,060	0,062	0,062	0,060	0,064	0,058	0,059	0,004	0,001
psbC_trnS-UGA	IGS	154		0,229	0,393	0,352	0,221	0,273		0,262	0,252	0,296	0,285	0,060	0,021
trnS-UGA	tRNA	93		0,023	0,023	0,011	0,047	0,035		0,035	0,035	0,047	0,032	0,012	0,004
trnS-UGA_psbZ	IGS	347		0,159	0,207	0,153	0,144	0,144		0,144	0,149	0,145	0,156	0,021	0,008
psbZ	CDS	189	0,096	0,078	0,078	0,072	0,078	0,049	0,061	0,061	0,067	0,067	0,071	0,013	0,004
psbZ_trnG-GCC	IGS	345	0,350			0,293			0,321	0,351	0,331	0,495	0,357	0,071	0,029
trnG-GCC	tRNA	71	0,000	0,052		0,000	0,052	0,191	0,017	0,017	0,000	0,000	0,037	0,062	0,021
trnG-GCC_trnfM-CAU	IGS	127	0,340	0,331		0,268	0,304		0,331	0,332	0,332	0,209	0,306	0,046	0,016
trnfM-CAU	tRNA	74	0,074	0,058	0,058	0,043	0,105	0,043	0,042	0,043	0,028	0,043	0,054	0,022	0,007
trnfM-CAU_rps14	IGS	161	0,142	0,177	0,110	0,142	0,120	0,131	0,121	0,089	0,110	0,132	0,128	0,024	0,008
rps14	CDS	303	0,048	0,154	0,048	0,051	0,137	0,091	0,099	0,076	0,051	0,052	0,081	0,039	0,012
rps14_psaB	IGS	132	0,276	0,292	0,277	0,277	0,277	0,245	0,245	0,245	0,277	0,215	0,263	0,024	0,008
psaB	CDS	2205	0,072	0,082	0,074	0,060	0,055	0,057	0,057	0,051	0,051	0,056	0,062	0,011	0,003
psaB_psaA	IGS	25	0,085	0,042	0,085	0,085	0,000	0,000	0,000	0,000	0,042	0,085	0,042	0,040	0,013
psaA	CDS	2253	0,059	0,055	0,062	0,055	0,053	0,048	0,048	0,048	0,056	0,059	0,054	0,005	0,002
psaA_ycf3	IGS	570	0,132		0,141	0,115	0,098	0,086		0,115	0,116	0,155	0,120	0,022	0,008
ycf3	CDS	153	0,055	0,053	0,057	0,048	0,059	0,064	0,064	0,042	0,073	0,046	0,056	0,009	0,003
ycf3_intron1	INTRON	727	0,135	0,086	0,137	0,109	0,093	0,080	0,101	0,086	0,109	0,086	0,102	0,021	0,007
ycf3	CDS	228	0,055	0,053	0,057	0,048	0,059	0,064	0,064	0,042	0,073	0,046	0,056	0,009	0,003
ycf3_intron2	INTRON	737	0,157	0,100	0,159	0,115	0,096	0,090	0,106	0,089	0,107	0,100	0,112	0,026	0,008
ycf3	CDS	126	0,055	0,053	0,057	0,048	0,059	0,064	0,064	0,042	0,073	0,046	0,056	0,009	0,003
ycf3_trnS-GGA	IGS	281		0,195	0,244	0,221	0,188	0,180	0,230	0,200	0,215	0,187	0,207	0,022	0,007
trnS-GGA	tRNA	88		0,024	0,037	0,024	0,024	0,037	0,024	0,049	0,037	0,037	0,032	0,009	0,003
trnS-GGA_psbI	IGS	191													
psbI	CDS	111	0,086	0,056	0,086	0,066	0,066	0,066	0,066	0,056	0,119	0,086	0,076	0,019	0,006
psbI_psbK	IGS	475	0,177	0,089	0,140	0,137	0,100	0,089	0,124	0,137	0,077	0,217	0,129	0,044	0,014
psbK	CDS	186	0,083	0,096	0,122	0,091	0,122	0,108	0,071	0,083	0,122	0,090	0,099	0,019	0,006
psbK_trnQ-UUG	IGS	473	0,167	0,287	0,142	0,134	0,307	0,118	0,118	0,094	0,176	0,134	0,168	0,072	0,023
trnQ-UUG	tRNA	72	0,000	0,058	0,000	0,014	0,059	0,014	0,029	0,044	0,000	0,000	0,022	0,024	0,008
trnQ-UUG_rps16	IGS	461	0,405		0,375	0,300					0,219	0,272	0,314	0,076	0,034
rps16	CDS	47				0,206					0,242	0,205	0,218	0,021	0,012
rps16_intron	INTRON	667				0,122					0,122	0,120	0,121	0,001	0,001
rps16	CDS	205				0,206					0,242	0,205	0,218	0,021	0,012
rps16_accD	IGS	407	0,290		0,256	0,244					0,242	0,205	0,247	0,030	0,014
accD	CDS	1497	0,167	0,252	0,126	0,115	0,304	0,136		0,263	0,108	0,093	0,174	0,078	0,026
accD_psaI	IGS	293	0,504	0,630	0,504	0,487		0,490		0,504	0,275	0,389	0,473	0,103	0,036
psaI	CDS	105	0,070	0,246	0,081	0,103		0,136	0,102	0,102	0,125	0,102	0,119	0,052	0,017
psaI_ycf4	IGS	201	0,201	0,355	0,138						0,052		0,186	0,128	0,064

ycf4	CDS	555	0,330		0,334		0,800				0,254	0,308	0,405	0,223	0,100
ycf4_cemA	IGS	317	0,333	0,460	0,333		0,423				0,350	0,244	0,357	0,076	0,031
cemA	CDS	690	0,135	0,145	0,128	0,106	0,154	0,117	0,109	0,102	0,097	0,089	0,118	0,022	0,007
cemA_petA	IGS	172	0,254	0,116	0,208	0,256	0,106	0,078	0,097	0,088	0,127	0,117	0,145	0,068	0,022
petA	CDS	963	0,096	0,088	0,104	0,084	0,086	0,087	0,081	0,080	0,084	0,083	0,087	0,007	0,002
petA_psbJ	IGS	1173	0,236	0,231	0,257	0,293		0,176	0,189	0,190	0,285	0,229	0,232	0,042	0,014
psbJ	CDS	123	0,034	0,060	0,033	0,017	0,051	0,033	0,051	0,060	0,033	0,008	0,038	0,017	0,005
psbJ_psbL	IGS	124	0,114	0,074	0,074	0,094	0,103	0,064	0,073	0,054	0,064	0,084	0,080	0,019	0,006
psbL	CDS	117	0,044	0,063	0,053	0,035	0,035	0,035	0,035	0,026	0,063	0,044	0,043	0,013	0,004
psbL_psbF	IGS	22	0,100	0,048	0,100	0,100	0,000	0,000	0,047	0,000	0,000	0,000	0,040	0,046	0,015
psbF	CDS	120	0,026	0,008	0,044	0,008	0,008	0,017	0,052	0,017	0,017	0,034	0,023	0,015	0,005
psbF_psbE	IGS	9	0,122	0,122	0,122	0,122	0,122	0,122	0,122	0,122	0,122	0,122	0,122	0,000	0,000
psbE	CDS	252	0,037	0,033	0,037	0,037	0,033	0,024	0,029	0,020	0,024	0,033	0,031	0,006	0,002
psbE_petL	IGS	1415	0,121		0,121	0,107		0,145	0,136	0,136	0,145	0,164	0,134	0,018	0,006
petL	CDS	96	0,114	0,125	0,112	0,078	0,125	0,101	0,126	0,113	0,089	0,067	0,105	0,021	0,007
petL_petG	IGS	163	0,261	0,216	0,272	0,238	0,216	0,183	0,184	0,143	0,238	0,186	0,214	0,040	0,013
petG	CDS	114	0,093	0,083	0,074	0,074	0,083	0,074	0,064	0,074	0,036	0,074	0,073	0,015	0,005
petG_trnW-CCA	IGS	136	0,346	0,313	0,327	0,355	0,346	0,258	0,259	0,244	0,182	0,197	0,283	0,064	0,020
trnW-CCA	tRNA	74	0,042	0,014	0,014	0,000	0,000	0,000	0,014	0,000	0,000	0,000	0,008	0,014	0,004
trnW-CCA_trnP-UGG	IGS	116	0,284	0,207	0,289	0,246	0,207	0,171	0,264	0,207	0,349	0,226	0,245	0,053	0,017
trnP-UGG	tRNA	74	0,000	0,014	0,000	0,000	0,014	0,014	0,014	0,014	0,014	0,000	0,008	0,007	0,002
trnP-UGG_psaJ	IGS	438	0,329	0,307	0,261	0,274	0,348	0,201	0,278	0,277	0,248	0,164	0,269	0,056	0,018
psaJ	CDS	129	0,112	0,113	0,094	0,094	0,103	0,085	0,094	0,114	0,076	0,094	0,098	0,013	0,004
psaJ_rpl33	IGS	513		0,305	0,311	0,207	0,285	0,246	0,234	0,271	0,292	0,253	0,267	0,035	0,012
rpl33	CDS	204		0,108		0,091	0,091	0,092	0,086	0,109	0,098	0,063	0,092	0,014	0,005
rpl33_rps18	IGS	220		0,240	0,259	0,194	0,223	0,274		0,239	0,243	0,258	0,241	0,024	0,009
rps18	CDS	330	0,063	0,201	0,072	0,072	0,223	0,126		0,152	0,072	0,044	0,114	0,065	0,022
rps18_rpl20	IGS	232	0,327	0,293	0,369	0,256	0,364	0,220		0,159	0,190	0,256	0,270	0,074	0,025
rpl20	CDS	357	0,142	0,164	0,146	0,131	0,149	0,134	0,207	0,120	0,120	0,132	0,144	0,026	0,008
rpl20_rps12	IGS	763	0,174	0,133	0,199	0,138	0,139	0,105	0,131	0,115	0,125	0,160	0,142	0,028	0,009
rps12	CDS	114	0,042	0,068	0,039	0,031	0,080	0,031	0,045	0,039	0,025	0,039	0,044	0,017	0,005
rps12_clpP	IGS	188	0,282	0,376	0,269	0,270	0,421	0,295		0,362	0,246	0,258	0,309	0,062	0,021
clpP	CDS	228	0,129	0,238	0,139	0,089	0,243	0,195	0,272	0,214	0,073	0,125	0,172	0,070	0,022
clpP_intron1	INTRON	655	0,321	0,267	0,317	0,318	0,246	0,241	0,236	0,299	0,189	0,267	0,270	0,044	0,014
clpP	CDS	292	0,129	0,238	0,139	0,089	0,243	0,195	0,272	0,214	0,073	0,125	0,172	0,070	0,022
clpP_intron2	INTRON	739	0,247		0,242	0,205					0,244	0,213	0,230	0,020	0,009
clpP	CDS	71	0,129	0,238	0,139	0,089	0,243	0,195	0,272	0,214	0,073	0,125	0,172	0,070	0,022
clpP_psbB	IGS	482	0,227		0,222	0,194		0,206		0,175	0,164	0,214	0,200	0,024	0,009
psbB	CDS	1530	0,077	0,060	0,072	0,072	0,059	0,060	0,067	0,054	0,054	0,057	0,063	0,008	0,003
psbB_psbT	IGS	182	0,209	0,195	0,250	0,182	0,195	0,182	0,250	0,195	0,170	0,195	0,202	0,027	0,009
psbT	CDS	108	0,073	0,073	0,083	0,062	0,062	0,051	0,062	0,040	0,063	0,041	0,061	0,014	0,004
psbT_psbN	IGS	60	0,153	0,054	0,174	0,112	0,072	0,154	0,112	0,073	0,073	0,132	0,111	0,042	0,013
psbN	CDS	132	0,039	0,031	0,048	0,047	0,031	0,031	0,072	0,047	0,031	0,023	0,040	0,014	0,004

psbN_psbH	IGS	109	0,113	0,100	0,113	0,100	0,100	0,074	0,127	0,073	0,061	0,036	0,090	0,028	0,009
psbH	CDS	222	0,057	0,082	0,057	0,052	0,082	0,072	0,109	0,078	0,098	0,058	0,075	0,019	0,006
psbH_petB	IGS	135	0,198	0,133	0,211	0,222	0,144	0,144	0,154	0,123	0,166	0,199	0,169	0,035	0,011
petB	CDS	6	0,065	0,065	0,067	0,060	0,063	0,068	0,051	0,051	0,066	0,066	0,062	0,006	0,002
petB_intron	INTRON	843	0,164	0,139	0,170	0,135	0,137	0,129	0,127	0,131	0,148	0,146	0,143	0,015	0,005
petB	CDS	642	0,065	0,065	0,067	0,060	0,063	0,068	0,051	0,051	0,066	0,066	0,062	0,006	0,002
petB_petD	IGS	214	0,218	0,211	0,219	0,158	0,211	0,174	0,220	0,243	0,159	0,204	0,202	0,029	0,009
petD	CDS	8	0,061	0,086	0,063	0,056	0,093	0,058	0,084	0,077	0,070	0,072	0,072	0,013	0,004
petD_intron	INTRON	743	0,257	0,227	0,253	0,213	0,231	0,247	0,239	0,196	0,178	0,201	0,224	0,027	0,008
petD	CDS	475	0,061	0,086	0,063	0,056	0,093	0,058	0,084	0,077	0,070	0,072	0,072	0,013	0,004
petD_rpoA	IGS	213	0,354	0,214	0,319	0,194	0,233	0,276		0,176	0,148	0,193	0,234	0,069	0,023
rpoA	CDS	996	0,129	0,099	0,122	0,086	0,105	0,098	0,121	0,083	0,098	0,089	0,103	0,016	0,005
rpoA_rps11	IGS	77	0,246	0,246	0,316	0,203	0,224	0,202	0,203	0,203	0,203	0,246	0,229	0,037	0,012
rps11	CDS	417	0,109	0,100	0,100	0,075	0,131	0,086	0,152	0,077	0,091	0,053	0,097	0,028	0,009
rps11_rpl36	IGS	295	0,455	0,358	0,388	0,358	0,343	0,274	0,261	0,287	0,261	0,303	0,329	0,063	0,020
rpl36	CDS	114	0,114	0,104	0,135	0,094	0,114	0,105	0,085	0,105	0,094	0,127	0,108	0,015	0,005
rpl36_rps8	IGS	453	0,338	0,434	0,398	0,300	0,511	0,330	0,305	0,338	0,333	0,340	0,363	0,066	0,021
rps8	CDS	405	0,140	0,151	0,140	0,093	0,241	0,144	0,126	0,126	0,082	0,108	0,135	0,044	0,014
rps8_rpl14	IGS	267		0,369		0,324	0,502	0,371	0,342	0,369	0,202	0,258	0,342	0,089	0,031
rpl14	CDS	369	0,108	0,124	0,105	0,086	0,174	0,102	0,108	0,105	0,081	0,093	0,108	0,026	0,008
rpl14_rpl16	IGS	130	0,292	0,239	0,278	0,212	0,271	0,166	0,225	0,194	0,224	0,215	0,232	0,039	0,012
rpl16	CDS	399	0,100	0,097	0,086	0,072	0,120	0,103	0,105	0,105	0,078	0,059	0,092	0,018	0,006
rpl16_intron	INTRON	1155	0,221	0,183	0,225	0,188	0,212	0,167		0,195	0,186	0,182	0,195	0,020	0,007
rpl16	CDS	9	0,100	0,097	0,086	0,072	0,120	0,103	0,105	0,105	0,078	0,059	0,092	0,018	0,006
rpl16_rps3	IGS	142	0,374	0,262	0,326	0,282	0,328	0,183	0,405	0,244	0,230	0,358	0,299	0,071	0,022
rps3	CDS	642	0,128	0,176	0,116	0,098	0,235	0,099	0,109	0,114	0,083	0,089	0,125	0,047	0,015
rps3_rps19	IGS	269		0,532		0,351	0,548	0,352	0,316	0,306	0,305	0,364	0,384	0,099	0,035
rps19	CDS	279	0,064	0,103	0,064	0,060	0,139	0,090	0,107	0,069	0,090	0,043	0,083	0,028	0,009
rps19_rpl2	IGS	64	0,148	0,149	0,197	0,148	0,197	0,276	0,224	0,149	0,125	0,103	0,172	0,052	0,016
rpl2	CDS	435	0,038	0,062	0,041	0,032	0,064	0,073	0,070	0,057	0,043	0,021	0,050	0,018	0,006
rpl2_intron	INTRON	672	0,033	0,079	0,031	0,033	0,067	0,048	0,062	0,060	0,038	0,043	0,049	0,017	0,005
rpl2	CDS	390	0,038	0,062	0,041	0,032	0,064	0,073	0,070	0,057	0,043	0,021	0,050	0,018	0,006
rpl2_rpl23	IGS	18	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
rpl23	CDS	282	0,026		0,022	0,034		0,053	0,077	0,085	0,026	0,011	0,042	0,027	0,010
rpl23_trnI-CAU	IGS	450	0,053			0,038	0,457	0,120		0,111	0,038		0,136	0,161	0,066
trnI-CAU	tRNA	74	0,000	0,028	0,000	0,000	0,057	0,028	0,028	0,014	0,000	0,000	0,015	0,019	0,006
trnI-CAU_ycf2	IGS	100	0,000	0,147	0,738	0,000		0,239	0,147	0,108	0,034	0,627	0,227	0,271	0,090
ycf2	CDS	6924	0,059	0,170	0,060	0,050	0,183	0,131	0,148	0,159	0,042	0,047	0,105	0,058	0,018
ycf2_trnL-CAA	IGS	580	0,046		0,052	0,045	0,129	0,168		0,122	0,079		0,092	0,049	0,018
trnL-CAA	tRNA	81	0,000	0,025	0,000	0,000	0,012	0,025	0,012	0,012	0,013	0,000	0,010	0,010	0,003
trnL-CAA_ndhB	IGS	578	0,073		0,062	0,054		0,135		0,136	0,041	0,054	0,079	0,040	0,015
ndhB	CDS	756	0,028	0,047	0,026	0,025	0,059	0,046		0,046	0,025	0,026	0,036	0,013	0,004
ndhB_intron	INTRON	675	0,039	0,058	0,040	0,034	0,068	0,051		0,059	0,032	0,026	0,045	0,014	0,005

ndhB	CDS	747	0,028	0,047	0,026	0,025	0,059	0,046		0,046	0,025	0,026	0,036	0,013	0,004
ndhB_rps7	IGS	349	0,085	0,257	0,085	0,085	0,288	0,157	0,202	0,203	0,041	0,085	0,149	0,085	0,027
rps7	CDS	468	0,024	0,067	0,022	0,017	0,126	0,042	0,053	0,042	0,020	0,019	0,043	0,034	0,011
rps7_rps12	IGS	65	0,000	0,170	0,025	0,000	0,168	0,168	0,138	0,174	0,053	0,137	0,103	0,075	0,024
rps12	CDS	26	0,042	0,068	0,039	0,031	0,080	0,031	0,045	0,039	0,025	0,039	0,044	0,017	0,005
rps12_intron	INTRON	539	0,027		0,031	0,031					0,023	0,027	0,028	0,003	0,001
rps12	CDS	232	0,042	0,068	0,039	0,031	0,080	0,031	0,045	0,039	0,025	0,039	0,044	0,017	0,005
rps12_trnV-GAC	IGS	1869	0,074	0,138	0,079	0,052		0,085	0,123	0,091	0,052	0,048	0,082	0,031	0,010
trnV-GAC	tRNA	72	0,090	0,074	0,090	0,090	0,074	0,074	0,074	0,107	0,074	0,074	0,082	0,012	0,004
trnV-GAC_rrn16	IGS	225	0,094	0,222	0,100	0,094	0,328	0,136	0,208	0,181	0,082	0,094	0,154	0,080	0,025
rrn16	rRNA	1491	0,005	0,031	0,005	0,005	0,042	0,023	0,028	0,023	0,006	0,005	0,017	0,014	0,004
rrn16_trnI-GAU	IGS	293	0,062	0,200	0,087	0,054	0,221	0,157		0,242	0,075	0,054	0,128	0,077	0,026
trnI-GAU	tRNA	38	0,000	0,000	0,000	0,000	0,000	0,014		0,000	0,000	0,000	0,002	0,005	0,002
trnI-GAU_intron	INTRON	953	0,030	0,051	0,027	0,022	0,064	0,047		0,034	0,022	0,022	0,035	0,015	0,005
trnI-GAU	tRNA	35	0,000	0,000	0,000	0,000	0,000	0,014		0,000	0,000	0,000	0,002	0,005	0,002
trnI-GAU_trnA-UGC	IGS	64	0,136	0,092	0,136	0,114	0,092	0,073		0,073	0,054	0,073	0,094	0,029	0,010
trnA-UGC	tRNA	38	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
trnA-UGC_intron	INTRON	821	0,019	0,058	0,019	0,019	0,058	0,058	0,366	0,037	0,019	0,019	0,067	0,107	0,034
trnA-UGC	tRNA	35	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
trnA-UGC_rrn23	IGS	157	0,023	0,141	0,023	0,023	0,072	0,200	0,047	0,085	0,024	0,061	0,070	0,059	0,019
rrn23	rRNA	2812	0,019	0,035	0,019	0,018	0,041	0,029	0,038	0,026	0,012	0,015	0,025	0,010	0,003
rrn23_rrn4.5	IGS	101	0,028	0,155	0,028	0,028	0,103	0,042	0,073	0,028	0,073	0,042	0,060	0,042	0,013
rrn4.5	rRNA	104	0,040	0,051	0,040	0,030	0,061	0,030	0,050	0,030	0,020	0,010	0,036	0,015	0,005
rrn4.5_rrn5	IGS	227	0,094	0,155	0,101	0,087	0,148	0,133	0,126	0,113	0,061	0,061	0,108	0,033	0,010
rrn5	rRNA	121	0,000	0,017	0,000	0,000	0,008	0,008	0,008	0,017	0,000	0,025	0,008	0,009	0,003
rrn5_trnR-ACG	IGS	255	0,096	0,167	0,076	0,050	0,191	0,176	0,190	0,190	0,103	0,089	0,133	0,055	0,017
trnR-ACG	tRNA	74	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
trnR-ACG_trnN-GUU	IGS	575	0,117	0,215	0,135	0,116	0,279	0,281	0,192	0,235	0,116	0,117	0,180	0,069	0,022
trnN-GUU	tRNA	72	0,000	0,075	0,000	0,000	0,044	0,059	0,044	0,044	0,029	0,000	0,029	0,028	0,009
trnN-GUU_ycf1	IGS	408	0,128		0,119	0,085		0,245		0,247	0,102	0,080	0,144	0,072	0,027
ycf1	CDS	5298	0,170	0,316	0,146	0,109	0,290	0,285		0,228	0,141	0,126	0,201	0,080	0,027
ycf1_rps15	IGS	449		0,378	0,357	0,347	0,277	0,330		0,302	0,333	0,349	0,334	0,032	0,011
rps15	CDS	276	0,158	0,172	0,154	0,126	0,197	0,185	0,125	0,140	0,098	0,103	0,146	0,033	0,011
rps15_ndhH	IGS	114	0,136	0,136	0,136	0,124	0,148	0,124	0,148	0,136	0,136	0,112	0,134	0,011	0,004
ndhH	CDS	1182	0,099	0,085	0,100	0,093	0,082	0,077	0,084	0,083	0,081	0,078	0,086	0,008	0,003
ndhH_ndhA	IGS	1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
ndhA	CDS	552	0,113	0,093	0,111	0,093	0,098	0,066	0,083	0,075	0,076	0,084	0,089	0,015	0,005
ndhA_intron	INTRON	1171	0,253	0,205	0,244	0,204	0,206	0,210	0,198	0,205	0,209	0,217	0,215	0,019	0,006
ndhA	CDS	540	0,113	0,093	0,111	0,093	0,098	0,066	0,083	0,075	0,076	0,084	0,089	0,015	0,005
ndhA_ndhI	IGS	79	0,185	0,159	0,159	0,110	0,185	0,135	0,135	0,110	0,064	0,087	0,133	0,040	0,013
ndhI	CDS	486	0,113	0,061	0,123	0,101	0,061	0,072	0,070	0,056	0,079	0,077	0,081	0,023	0,007
ndhI_ndhG	IGS	374	0,197	0,158	0,217	0,269	0,168	0,168	0,168	0,279	0,228	0,198	0,205	0,043	0,014
ndhG	CDS	531	0,116	0,079	0,116	0,088	0,079	0,069	0,073	0,065	0,080	0,082	0,085	0,018	0,006

ndhG_ndhE	IGS	247	0,180		0,163	0,147	0,115	0,164	0,116	0,117	0,181	0,115	0,144	0,029	0,010
ndhE	CDS	306	0,072	0,044	0,069	0,065	0,040	0,051	0,054	0,040	0,062	0,058	0,056	0,012	0,004
ndhE_psaC	IGS	268	0,209	0,244	0,195	0,176	0,237	0,195	0,175	0,217	0,163	0,170	0,198	0,028	0,009
psaC	CDS	246	0,073	0,069	0,060	0,068	0,073	0,069	0,091	0,078	0,064	0,064	0,071	0,009	0,003
psaC_ndhD	IGS	113	0,247	0,105	0,245	0,193	0,107	0,057	0,106	0,093	0,131	0,149	0,143	0,065	0,021
ndhD	CDS	1515	0,114	0,111	0,115	0,098	0,098	0,091	0,090	0,089	0,095	0,095	0,100	0,010	0,003
ndhD_ccsA	IGS	296	0,173		0,194	0,163	0,218	0,143	0,226	0,194	0,226	0,174	0,190	0,029	0,010
ccsA	CDS	963	0,144	0,113	0,145	0,120	0,117	0,111	0,111	0,110	0,111	0,103	0,118	0,015	0,005
ccsA_trnL-UAG	IGS	103	0,317	0,174	0,336	0,258	0,229	0,192	0,140	0,140	0,249	0,253	0,229	0,068	0,021
trnL-UAG	CDS	80	0,025	0,025	0,039	0,025	0,025	0,025	0,025	0,025	0,025	0,025	0,027	0,004	0,001
trnL-UAG_rpl32	IGS	314	0,347	0,328	0,329	0,320		0,303		0,293	0,311	0,375	0,326	0,026	0,009
rpl32	CDS	147	0,212	0,223	0,202	0,167	0,254	0,175		0,140	0,184	0,149	0,189	0,037	0,012
rpl32_ndhF	IGS	476	0,295	0,210	0,282	0,228	0,209	0,198		0,215	0,257	0,239	0,237	0,034	0,011
ndhF	CDS	2241	0,161	0,116	0,165	0,129	0,122	0,121	0,110	0,109	0,120	0,132	0,129	0,019	0,006

Table S5. Synonymous mutation rate of protein-coding sequences of the *L. luteus* plastome and their homologs in ten other Fabaceae. The synonymous mutation rate has been calculated using the yn00 method with PAML program.

Region	*Phaseolus vulgaris*	*Pisum sativum*	*Vigna radiata*	*Glycine max*	*Lathyrus sativus*	*Cicer arietinum*	*Trifolium subterraneum*	*Medicago truncatula*	*Lotus japonicus*	*Millettia pinnata*	*mean*	*standard deviation*	*standard error*
accD	0,4137	0,7669	0,4157	0,4285	1,0941	0,49		0,6537	0,3673	0,2108	0,538	0,264	0,088
atpA	0,3211	0,3034	0,3283	0,2862	0,3552	0,3607	0,3126	0,3378	0,3324	0,3019	0,324	0,024	0,008
atpB	0,3729	0,334	0,3762	0,323	0,3685	0,3316	0,3492	0,3592	0,3191	0,3042	0,344	0,025	0,008
atpE	0,4028	0,3032	0,434	0,3571	0,3172	0,3186	0,2794	0,3029	0,3528	0,3793	0,345	0,049	0,016
atpF	0,2992	0,3224	0,2552	0,212	0,3001	0,2959	0,306	0,2897	0,2345	0,2071	0,272	0,042	0,013
atpH	0,3195	0,4037	0,2879	0,36	0,4027	0,4531	0,413	0,4222	0,2702	0,2236	0,356	0,076	0,024
atpI	0,1672	0,224	0,1632	0,1762	0,2249	0,2382	0,2376	0,252	0,3518	0,1518	0,219	0,059	0,019
ccsA	0,307	0,2652	0,3149	0,2836	0,2778	0,2889	0,251	0,2851	0,2767	0,1953	0,275	0,033	0,011
cemA	0,3131	0,3133	0,3295	0,2613	0,2723	0,3458	0,2596	0,3027	0,1753	0,2696	0,284	0,049	0,015
clpP	0,4558	0,3362	0,4377	0,394	0,3457	0,4532	0,6064	0,4561	0,2723	0,3743	0,413	0,092	0,029
matK	0,2914	0,2454	0,3137	0,2574	0,2242	0,2001	0,2215	0,1979	0,2155	0,1832	0,235	0,042	0,013
ndhA	0,3625	0,3079	0,3566	0,2973	0,336	0,2181	0,2669	0,2532	0,2585	0,2773	0,293	0,047	0,015
ndhB	0,0616	0,0988	0,0633	0,061	0,1314	0,1062		0,1102	0,047	0,0563	0,082	0,030	0,010
ndhC	0,3236	0,2984	0,3876	0,2616	0,3122	0,3045	0,3024	0,3504	0,2535	0,2274	0,302	0,047	0,015
ndhD	0,3056	0,3026	0,2918	0,2539	0,2774	0,2411	0,2464	0,2469	0,2247	0,2397	0,263	0,029	0,009
ndhE	0,1918	0,1269	0,1862	0,2395	0,1073	0,1596	0,164	0,0918	0,2734	0,1421	0,168	0,057	0,018
ndhF	0,5535	0,4034	0,6156	0,4708	0,4176	0,4354	0,3508	0,3743	0,468	0,448	0,454	0,080	0,025
ndhG	0,3181	0,2315	0,3295	0,3303	0,2596	0,215	0,2134	0,1809	0,2012	0,2454	0,252	0,055	0,017
ndhH	0,3954	0,3562	0,4483	0,4045	0,3632	0,3591	0,41	0,4445	0,4237	0,3339	0,394	0,039	0,012
ndhI	0,4636	0,2324	0,5821	0,3941	0,2207	0,3009	0,3236	0,2288	0,4283	0,296	0,347	0,118	0,037
ndhJ	0,2319	0,2101	0,239	0,2178	0,214	0,2147	0,2042	0,188	0,2132	0,2494	0,218	0,018	0,006
ndhK	0,2513	0,2293	0,2643	0,2535	0,2393	0,364		0,2146	0,2838	0,2188	0,258	0,046	0,015
petA	0,3541	0,3659	0,4236	0,2912	0,3264	0,3676	0,3632	0,3915	0,2797	0,3422	0,351	0,043	0,014
petB	0,3985	0,3399	0,4272	0,3128	0,3351	0,3568	0,249	0,2451	0,3725	0,3111	0,335	0,059	0,019
petD	0,2446	0,3129	0,2645	0,2037	0,3642	0,2151	0,3135	0,282	0,2509	0,2543	0,271	0,049	0,015
petG	0,5208	0,5069	0,2294	0,3049	0,5069	0,4666	0,2244	0,2608	0,1386	0,2118	0,337	0,147	0,046
petL	0,2483	0,1882	0,3121	0,0896	0,2019	0,1928	0,3031	0,259	0,1388	0,1593	0,209	0,071	0,023
petN	0,1287	0,0572	0,0563	0,1272	0,1306	0,0588	0,3708	0,0584	0,1157	0,0565	0,116	0,096	0,030
psaA	0,226	0,2272	0,2414	0,2161	0,2249	0,1955	0,1934	0,188	0,2243	0,2379	0,217	0,019	0,006
psaB	0,3107	0,3375	0,3436	0,2529	0,2341	0,2695	0,2626	0,2169	0,2865	0,2306	0,274	0,044	0,014
psaC	0,3609	0,3828	0,3332	0,3947	0,4255	0,3961	0,576	0,4192	0,3916	0,348	0,403	0,068	0,021
psaI	0,3545	0,652	0,4567	0,6311		0,5545	0,3604	0,3524	0,4114	0,4205	0,466	0,118	0,039
psaJ	0,2894	0,2169	0,2462	0,2281	0,1847	0,157	0,1928	0,2132	0,1852	0,2277	0,214	0,037	0,012
psbA	0,2922	0,1382	0,2887	0,2485	0,1353	0,1619	0,2364	0,156	0,171	0,2307	0,206	0,060	0,019

psbB	0,3292	0,2247	0,317	0,2868	0,2007	0,2552	0,2553	0,2014	0,2097	0,2149	0,249	0,048	0,015
psbC	0,2405	0,2535	0,2599	0,2317	0,257	0,2745	0,2605	0,2255	0,2835	0,2519	0,254	0,018	0,006
psbD	0,246	0,1619	0,1843	0,1428	0,1692	0,182	0,1832	0,1636	0,1594	0,1556	0,175	0,028	0,009
psbE	0,1463	0,1193	0,1499	0,1626	0,1193	0,0962	0,1051	0,0703	0,1299	0,1212	0,122	0,027	0,009
psbF	0,1025	0,0309	0,1418	0,0317	0,0309	0,0627	0,1845	0,0311	0,0318	0,1399	0,079	0,059	0,019
psbH	0,169	0,273	0,1659	0,1756	0,2763	0,2389	0,4598	0,2535	0,3684	0,2539	0,263	0,092	0,029
psbI	0,3438	0,1891	0,3351	0,2756	0,2566	0,2337	0,2362	0,1939	0,5298	0,4152	0,301	0,107	0,034
psbJ	0,1159	0,0738	0,1203	0,0367	0,0742	0,1135	0,1625	0,145	0,1219	0	0,096	0,050	0,016
psbK	0,1724	0,2871	0,2685	0,2127	0,2795	0,3178	0,2038	0,2208	0,2635	0,1495	0,238	0,054	0,017
psbL	0,1395	0,288	0,173	0,1338	0,1018	0,1338	0,1349	0,1004	0,2394	0,1709	0,162	0,060	0,019
psbM	0,2514	0,2274	0,2514	0,463	0,33	0,2899	0,2274	0,3321	0,4142	0,2187	0,301	0,084	0,027
psbN	0,1378	0,0715	0,1793	0,1383	0,0715	0,0751	0,1958	0,1456	0,125	0,0735	0,121	0,046	0,015
psbT	0,3794	0,2035	0,4538	0,2742	0,1671	0,1246	0,1365	0,0626	0,1593	0,0936	0,205	0,127	0,040
psbZ	0,3155	0,4309	0,2517	0,2485	0,3766	0,2423	0,2538	0,2675	0,291	0,3674	0,305	0,066	0,021
rbcL	0,3041	0,3668	0,3486	0,3319	0,3695	0,3951	0,4243	0,4349	0,3855	0,3078	0,367	0,045	0,014
rpl14	0,4763	0,5328	0,435	0,3521	0,7718	0,4227	0,377	0,4122	0,2966	0,3322	0,441	0,135	0,043
rpl16	0,594	0,2919	0,4941	0,375	0,3661	0,2963	0,3261	0,2816	0,3694	0,248	0,364	0,106	0,034
rpl20	0,4614	0,2557	0,409	0,3832	0,2337	0,2935	0,3856	0,2471	0,321	0,4196	0,341	0,081	0,026
rpl23	0,0258		0,0257	0,0604		0,1143	0,1487	0,1638	0,0168	0,0171	0,072	0,062	0,022
rpl2	0,1079	0,1642	0,1193	0,0999	0,1893	0,1629	0,1574	0,1993	0,1003	0,0653	0,137	0,044	0,014
rpl32	0,6147	1,0231	0,4735	0,2543	0,7495	0,4469		0,3334	0,6299	0,3534	0,542	0,241	0,080
rpl33		0,3949		0,2237	0,3467	0,2262	0,1697	0,2372	0,2133	0,2528	0,258	0,075	0,026
rpl36	0,4398	0,1714	0,6199	0,6724	0,2632	0,3307	0,3165	0,4336	0,4115	0,752	0,441	0,187	0,059
rpoA	0,3078	0,2436	0,3065	0,2243	0,1898	0,2916	0,3062	0,2253	0,2431	0,2678	0,261	0,042	0,013
rpoB	0,3163	0,2935	0,2915	0,2708	0,2868	0,2815	0,3019	0,2729	0,2964	0,2396	0,285	0,021	0,007
rpoC1	0,2769	0,2798	0,3008	0,2336	0,2967	0,2708	0,3156	0,308	0,312	0,1909	0,279	0,039	0,012
rpoC2	0,4623	0,3819	0,4889	0,4441	0,4172	0,422	0,3796	0,3672	0,4328	0,3948	0,419	0,039	0,012
rps11	0,4377	0,2981	0,4224	0,306	0,2839	0,2563	0,3624	0,2076	0,3662	0,1925	0,313	0,083	0,026
rps12	0,1729	0,2283	0,1769	0,1257	0,2478	0,0987	0,1567	0,1493	0,088	0,169	0,161	0,051	0,016
rps14	0,1773	0,341	0,1626	0,1736	0,1942	0,2031	0,1963	0,1603	0,1608	0,1142	0,188	0,059	0,019
rps15	0,3358	0,3136	0,2831	0,3814	0,3725	0,2869	0,3125	0,361	0,2507	0,1858	0,308	0,060	0,019
rps16				0,5403					0,9066	0,5544	0,667	0,208	0,120
rps18	0,1643	0,5152	0,1735	0,2308	0,8786	0,1833		0,1942	0,3068	0,1092	0,306	0,245	0,082
rps19	0,1775	0,2518	0,1775	0,1871	0,7514	0,295	0,567	0,2815	0,2817	0,1152	0,309	0,198	0,063
rps2	0,3178	0,408	0,3514	0,3516	0,3669	0,3127	0,3643	0,1658	0,3057	0,3088	0,325	0,065	0,020
rps3	0,4497	0,2403	0,3944	0,2967	0,2956	0,2556	0,2117	0,2331	0,3393	0,3728	0,309	0,078	0,025
rps4	0,334	0,2771	0,3471	0,2603	0,2697	0,2387	0,2786	0,2052	0,2647	0,2695	0,274	0,041	0,013
rps7	0,0689	0,1407	0,0682	0,0683	0,1479	0,1166	0,0819	0,0574	0,079	0,0807	0,091	0,032	0,010
rps8	0,4283	0,4564	0,4137	0,3425	0,6564	0,4034	0,3338	0,3602	0,1886	0,298	0,388	0,122	0,038
ycf1	0,2382	0,8125	0,4369	0,3902	0,8958	0,5512		0,537	0,3949	0,4028	0,518	0,212	0,071
ycf2	0,069	0,2677	0,07	0,0656	0,2763	0,1946	0,2292	0,2508	0,0564	0,0583	0,154	0,097	0,031
ycf3	0,1986	0,1946	0,2109	0,162	0,2493	0,3109	0,2439	0,1764	0,3378	0,1555	0,224	0,061	0,019
ycf4	0,6356		0,694		1,3108				0,2962	0,3563	0,659	0,403	0,180

Table S6. Non Synonymous mutation rate of protein coding sequences of the *L. luteus* plastome and their homologs in ten other Fabaceae plastomes. Non synonymous mutation rate has been calculated using the yn00 method with PAML program.

Region	*Phaseolus vulgaris*	*Pisum sativum*	*Vigna radiata*	*Glycine max*	*Lathyrus sativus*	*Cicer arietinum*	*Trifolium subterraneum*	*Medicago truncatula*	*Lotus japonicus*	*Millettia pinnata*	*mean*	*standard deviation*	*standard error*
accD	0,1865	0,3109	0,1487	0,1025	0,3216	0,1291		0,2719	0,0963	0,0848	0,184	0,094	0,031
atpA	0,0176	0,0276	0,0176	0,0134	0,0311	0,0215	0,0281	0,0242	0,0185	0,0197	0,022	0,006	0,002
atpB	0,0244	0,0363	0,0251	0,0185	0,0362	0,0275	0,0398	0,0287	0,0169	0,0182	0,027	0,008	0,003
atpE	0,0489	0,0888	0,0422	0,0346	0,0841	0,0393	0,0535	0,0606	0,0598	0,0423	0,055	0,018	0,006
atpF	0,0859	0,0804	0,0864	0,0628	0,0895	0,0724	0,0554	0,0707	0,0707	0,0786	0,075	0,011	0,003
atpH	0,0083	0,0142	0,0084	0,0081	0,0142	0,0223	0,0084	0,014	0,0083	0	0,011	0,006	0,002
atpI	0,0202	0,0182	0,0168	0,0235	0,0181	0,0201	0,0215	0,0173	0,0141	0,0225	0,019	0,003	0,001
ccsA	0,1098	0,0825	0,1094	0,085	0,0849	0,0727	0,0794	0,0728	0,075	0,0796	0,085	0,014	0,004
cemA	0,0916	0,1085	0,0818	0,0654	0,1279	0,0624	0,0697	0,0517	0,0695	0,0498	0,078	0,025	0,008
clpP	0,0623	0,2242	0,0768	0,0247	0,2279	0,1445	0,2147	0,1608	0,0206	0,067	0,122	0,082	0,026
matK	0,184	0,117	0,1803	0,1509	0,1123	0,1085	0,106	0,1024	0,1261	0,1295	0,132	0,030	0,010
ndhA	0,0587	0,0443	0,055	0,0458	0,0441	0,0313	0,038	0,0343	0,0375	0,0418	0,043	0,009	0,003
ndhB	0,0142	0,0267	0,0108	0,0102	0,0334	0,025		0,0226	0,0144	0,0132	0,019	0,008	0,003
ndhC	0,0297	0,0322	0,0296	0,0217	0,0264	0,0245	0,0288	0,0247	0,0359	0,0296	0,028	0,004	0,001
ndhD	0,0603	0,0582	0,0653	0,0519	0,0509	0,049	0,0455	0,0463	0,0565	0,0533	0,054	0,006	0,002
ndhE	0,0391	0,0192	0,035	0,0258	0,0193	0,0194	0,0236	0,0234	0,0203	0,031	0,026	0,007	0,002
ndhF	0,1133	0,0763	0,1127	0,0871	0,0893	0,0735	0,0714	0,068	0,0709	0,0867	0,085	0,017	0,005
ndhG	0,0728	0,0457	0,0705	0,0371	0,0394	0,0367	0,0417	0,039	0,0485	0,0479	0,048	0,013	0,004
ndhH	0,0394	0,0276	0,0352	0,0304	0,023	0,0243	0,0289	0,0224	0,0189	0,0267	0,028	0,006	0,002
ndhI	0,0445	0,0189	0,0444	0,0373	0,0191	0,0178	0,0176	0,011	0,0182	0,0272	0,026	0,012	0,004
ndhJ	0,0371	0,0456	0,0358	0,0243	0,035	0,0275	0,0266	0,0243	0,021	0,0188	0,030	0,008	0,003
ndhK	0,0442	0,0788	0,0438	0,0418	0,0723	0,068		0,0682	0,0906	0,0474	0,062	0,018	0,006

petA	0,0375	0,0318	0,0361	0,0308	0,0311	0,0277	0,0198	0,0168	0,0356	0,0247	0,029	0,007	0,002
petB	0,0058	0,004	0,0058	0,0079	0,0059	0,004	0,006	0,0089	0,004	0,012	0,006	0,003	0,001
petD	0,0138	0,0147	0,0137	0,014	0,0187	0,0114	0,0174	0,0202	0,0172	0,02	0,016	0,003	0,001
petG	0,0117	0,0114	0,0255	0,0123	0,0114	0	0,0127	0,0244	0	0,026	0,014	0,009	0,003
petL	0,0775	0,1107	0,0604	0,0769	0,1085	0,077	0,0778	0,0759	0,0764	0,0428	0,078	0,020	0,006
petN	0,0602	0,0451	0,061	0,0448	0,029	0,0294	0,0302	0,0448	0,0298	0,0454	0,042	0,012	0,004
psaA	0,0117	0,0115	0,014	0,0096	0,0059	0,0071	0,0077	0,0077	0,0073	0,0125	0,010	0,003	0,001
psaB	0,0144	0,0291	0,0134	0,0079	0,0129	0,0087	0,0111	0,0101	0,0068	0,0108	0,013	0,006	0,002
psaC	0,0053	0,0053	0	0	0,0053	0	0,0155	0,0106	0	0,0053	0,005	0,005	0,002
psaI	0,0127	0,1726	0,0126	0,0129		0,0654	0,0517	0,0523	0,0786	0,0513	0,057	0,050	0,017
psaJ	0,0353	0,0506	0,0235	0,0244	0,0506	0,0506	0,0501	0,0653	0,0231	0,0245	0,040	0,015	0,005
psbA	0,0031	0,0025	0,003	0,0024	0,0037	0,0024	0,0047	0,0037	0,0013	0,0037	0,003	0,001	0,000
psbB	0,0145	0,0147	0,0117	0,0133	0,0184	0,0154	0,0141	0,013	0,0104	0,0106	0,014	0,002	0,001
psbC	0,0046	0,0082	0,0073	0,0037	0,0093	0,0065	0,0047	0,0091	0,0047	0,0065	0,006	0,002	0,001
psbD	0,0025	0,0105	0,0026	0,0038	0,0064	0,005	0,0065	0,0039	0,0026	0,0026	0,005	0,003	0,001
psbE	0,0054	0,0104	0,0053	0	0,0104	0,0052	0,0053	0,0053	0	0,0103	0,006	0,004	0,001
psbF	0	0	0,0117	0	0	0	0,0117	0,012	0,0119	0	0,005	0,006	0,002
psbH	0,0297	0,0364	0,0361	0,0235	0,0362	0,0301	0,0418	0,0496	0,0471	0,0173	0,035	0,010	0,003
psbI	0	0,013	0	0	0,0127	0,013	0,0129	0,0129	0	0	0,006	0,007	0,002
psbJ	0,0109	0,0454	0,0109	0,0109	0,0335	0,011	0,0221	0,0227	0,0108	0,0111	0,019	0,012	0,004
psbK	0,0482	0,0435	0,0776	0,0469	0,0778	0,051	0,03	0,037	0,0807	0,0651	0,056	0,018	0,006
psbL	0,0123	0,0114	0,0124	0	0,0123	0	0	0	0,012	0	0,006	0,006	0,002
psbM	0,0244	0,0124	0,0244	0,0119	0,0122	0,0125	0,0124	0,0257	0,013	0,0125	0,016	0,006	0,002
psbN	0,0104	0,0203	0,0103	0,021	0,0203	0,0201	0,0416	0,0207	0	0	0,016	0,012	0,004
psbT	0,0125	0,056	0,0125	0,0271	0,0553	0,0259	0,0588	0,0586	0,0142	0,0334	0,035	0,020	0,006
psbZ	0,0558	0,0137	0,042	0,0348	0,0206	0,0069	0,0137	0,0205	0,0142	0,0135	0,024	0,016	0,005
rbcL	0,0125	0,0216	0,0192	0,0188	0,0206	0,0088	0,0213	0,021	0,0119	0,016	0,017	0,005	0,001
rpl14	0,039	0,0532	0,0392	0,0319	0,0805	0,039	0,054	0,0462	0,029	0,0365	0,045	0,015	0,005
rpl16	0,0219	0,051	0,0156	0,0126	0,0648	0,0542	0,0575	0,0456	0,019	0,0129	0,036	0,021	0,007
rpl20	0,0943	0,1467	0,1107	0,0788	0,1285	0,1038	0,1919	0,1004	0,0777	0,0714	0,110	0,037	0,012
rpl23	0,0256		0,0208	0,0258		0,0365	0,0577	0,0671	0,028	0,0091	0,034	0,019	0,007
rpl2	0,0177	0,0339	0,0194	0,0129	0,0322	0,0464	0,0392	0,0217	0,023	0,0099	0,026	0,012	0,004

rpl32	0,1515	0,1892	0,162	0,1517	0,1777	0,1333		0,1322	0,1165	0,1088	0,147	0,027	0,009
rpl33		0,0725		0,071	0,0594	0,0673	0,075	0,0892	0,0772	0,0318	0,068	0,017	0,006
rpl36	0,0587	0,0916	0,0592	0,0114	0,0842	0,0599	0,0353	0,0466	0,024	0,0461	0,052	0,025	0,008
rpoA	0,0982	0,0763	0,0889	0,0575	0,0903	0,0651	0,0904	0,0561	0,0709	0,0563	0,075	0,016	0,005
rpoB	0,0478	0,0429	0,0445	0,0387	0,0599	0,0397	0,0524	0,0386	0,0421	0,0425	0,045	0,007	0,002
rpoC1	0,0434	0,0516	0,0453	0,0308	0,072	0,0432	0,0562	0,0422	0,04	0,04	0,046	0,011	0,004
rpoC2	0,1088	0,0987	0,1121	0,0924	0,1095	0,1355	0,1063	0,0951	0,0916	0,0916	0,104	0,014	0,004
rps11	0,0457	0,0593	0,0422	0,0282	0,0969	0,0458	0,1078	0,0485	0,0316	0,019	0,053	0,029	0,009
rps12	0,0036	0,0239	0	0	0,0364	0,0074	0,0075	0,0037	0,0037	0	0,009	0,012	0,004
rps14	0,0191	0,116	0,0213	0,0198	0,1263	0,0663	0,079	0,0585	0,0222	0,0339	0,056	0,040	0,013
rps15	0,1174	0,1081	0,1214	0,0708	0,1491	0,1591	0,0829	0,0969	0,0569	0,0833	0,105	0,033	0,010
rps16				0,1809					0,1772	0,1503	0,169	0,017	0,010
rps18	0,0661	0,1665	0,0684	0,072	0,3225	0,1095		0,1713	0,0537	0,0486	0,120	0,089	0,030
rps19	0,0416	0,0792	0,0416	0,0365	0,0812	0,0595	0,047	0,0458	0,0559	0,0313	0,052	0,017	0,005
rps2	0,0636	0,0502	0,0604	0,0436	0,0659	0,0431	0,0816	0,0151	0,0443	0,0382	0,051	0,018	0,006
rps3	0,0746	0,179	0,0739	0,0658	0,2414	0,0663	0,0857	0,0879	0,0392	0,0447	0,096	0,064	0,020
rps4	0,0709	0,0763	0,0727	0,0579	0,0865	0,0615	0,0704	0,0554	0,0633	0,0796	0,069	0,010	0,003
rps7	0,0135	0,0504	0,0108	0,0054	0,1223	0,0247	0,0464	0,0384	0,0054	0,0054	0,032	0,036	0,011
rps8	0,0861	0,0898	0,0879	0,0464	0,1745	0,0884	0,0821	0,0758	0,0568	0,0702	0,086	0,034	0,011
ycf1	0,1742	0,4301	0,2689	0,2202	0,5031	0,4035		0,338	0,2085	0,2081	0,306	0,117	0,039
ycf2	0,051	0,1701	0,0518	0,042	0,1912	0,13	0,1458	0,1612	0,0384	0,0432	0,102	0,062	0,020
ycf3	0,0207	0,0107	0,0207	0,0184	0,0128	0,0126	0,0157	0,0105	0,0154	0,0157	0,015	0,004	0,001
ycf4	0,2945		0,2877		0,8447				0,249	0,3136	0,398	0,251	0,112

Table S7. Ka/Ks ratio between *L. luteus* and ten other Fabaceae plastomes. The ratio has been calculated using the yn00 method with PAML program.

Region	*Phaseolus vulgaris*	*Pisum sativum*	*Vigna radiata*	*Glycine max*	*Lathyrus sativus*	*Cicer arietinum*	*Trifolium subterraneum*	*Medicago truncatula*	*Lotus japonicus*	*Millettia pinnata*	*mean*	*standard deviation*	*standard error*
accD	0,4508	0,4054	0,3577	0,2392	0,2939	0,2635	-	0,4159	0,2622	0,4023	0,343	0,080	0,025
atpA	0,0548	0,0910	0,0536	0,0468	0,0876	0,0596	0,0899	0,0716	0,0557	0,0653	0,068	0,017	0,005
atpB	0,0654	0,1087	0,0667	0,0573	0,0982	0,0829	0,1140	0,0799	0,0530	0,0598	0,079	0,022	0,007
atpE	0,1214	0,2929	0,0972	0,0969	0,2651	0,1234	0,1915	0,2001	0,1695	0,1115	0,167	0,070	0,022
atpF	0,2871	0,2494	0,3386	0,2962	0,2982	0,2447	0,1810	0,2440	0,3015	0,3795	0,282	0,055	0,018
atpH	0,0260	0,0352	0,0292	0,0225	0,0353	0,0492	0,0203	0,0332	0,0307	0,0000	0,028	0,013	0,004
atpI	0,1208	0,0813	0,1029	0,1334	0,0805	0,0844	0,0905	0,0687	0,0401	0,1482	0,095	0,032	0,010
ccsA	0,3577	0,3111	0,3474	0,2997	0,3056	0,2516	0,3163	0,2553	0,2711	0,4076	0,312	0,049	0,015
cemA	0,2926	0,3463	0,2483	0,2503	0,4697	0,1805	0,2685	0,1708	0,3965	0,1847	0,281	0,098	0,031
clpP	0,1367	0,6669	0,1755	0,0627	0,6592	0,3188	0,3541	0,3526	0,0757	0,1790	0,298	0,219	0,069
matK	0,6314	0,4768	0,5748	0,5862	0,5009	0,5422	0,4786	0,5174	0,5852	0,7069	0,560	0,073	0,023
ndhA	0,1619	0,1439	0,1542	0,1541	0,1313	0,1435	0,1424	0,1355	0,1451	0,1507	0,146	0,009	0,003
ndhB	0,2305	0,2702	0,1706	0,1672	0,2542	0,2354	-	0,2051	0,3064	0,2345	0,230	0,045	0,014
ndhC	0,0918	0,1079	0,0764	0,0830	0,0846	0,0805	0,0952	0,0705	0,1416	0,1302	0,096	0,024	0,007
ndhD	0,1973	0,1923	0,2238	0,2044	0,1835	0,2032	0,1847	0,1875	0,2514	0,2224	0,205	0,022	0,007
ndhE	0,2039	0,1513	0,1880	0,1077	0,1799	0,1216	0,1439	0,2549	0,0743	0,2182	0,164	0,055	0,017
ndhF	0,2047	0,1891	0,1831	0,1850	0,2138	0,1688	0,2035	0,1817	0,1515	0,1935	0,187	0,018	0,006
ndhG	0,2289	0,1974	0,2140	0,1123	0,1518	0,1707	0,1954	0,2156	0,2411	0,1952	0,192	0,038	0,012
ndhH	0,0996	0,0775	0,0785	0,0752	0,0633	0,0677	0,0705	0,0504	0,0446	0,0800	0,071	0,016	0,005
ndhI	0,0960	0,0813	0,0763	0,0946	0,0865	0,0592	0,0544	0,0481	0,0425	0,0919	0,073	0,020	0,006
ndhJ	0,1600	0,2170	0,1498	0,1116	0,1636	0,1281	0,1303	0,1293	0,0985	0,0754	0,136	0,039	0,012
ndhK	0,1759	0,3437	0,1657	0,1649	0,3021	0,1868	-	0,3178	0,3192	0,2166	0,244	0,075	0,024
petA	0,1059	0,0869	0,0852	0,1058	0,0953	0,0754	0,0545	0,0429	0,1273	0,0722	0,085	0,025	0,008
petB	0,0146	0,0118	0,0136	0,0253	0,0176	0,0112	0,0241	0,0363	0,0107	0,0386	0,020	0,010	0,003
petD	0,0564	0,0470	0,0518	0,0687	0,0513	0,0530	0,0555	0,0716	0,0686	0,0786	0,060	0,011	0,003
petG	0,0225	0,0225	0,1112	0,0403	0,0225	0,0000	0,0566	0,0936	0,0000	0,1228	0,049	0,045	0,014
petL	0,3121	0,5882	0,1935	0,8583	0,5374	0,3994	0,2567	0,2931	0,5504	0,2687	0,426	0,205	0,065
petN	0,4678	0,7885	1,0835	0,3522	0,2221	0,5000	0,0814	0,7671	0,2576	0,8035	0,532	0,318	0,101
psaA	0,0518	0,0506	0,0580	0,0444	0,0262	0,0363	0,0398	0,0410	0,0325	0,0525	0,043	0,010	0,003
psaB	0,0463	0,0862	0,0390	0,0312	0,0551	0,0323	0,0423	0,0466	0,0237	0,0468	0,045	0,017	0,005
psaC	0,0147	0,0138	0,0000	0,0000	0,0125	0,0000	0,0269	0,0253	0,0000	0,0152	0,011	0,010	0,003
psaI	0,0358	0,2647	0,0276	0,0204	-	0,1179	0,1435	0,1484	0,1911	0,1220	0,119	0,081	0,026
psaJ	0,1220	0,2333	0,0955	0,1070	0,2740	0,3223	0,2599	0,3063	0,1247	0,1076	0,195	0,092	0,029
psbA	0,0106	0,0181	0,0104	0,0097	0,0273	0,0148	0,0199	0,0237	0,0076	0,0160	0,016	0,006	0,002

psbB	0,0440	0,0654	0,0369	0,0464	0,0917	0,0603	0,0552	0,0645	0,0496	0,0493	0,056	0,015	0,005
psbC	0,0191	0,0323	0,0281	0,0160	0,0362	0,0237	0,0180	0,0404	0,0166	0,0258	0,026	0,009	0,003
psbD	0,0102	0,0649	0,0141	0,0266	0,0378	0,0275	0,0355	0,0238	0,0163	0,0167	0,027	0,016	0,005
psbE	0,0369	0,0872	0,0354	0,0000	0,0872	0,0541	0,0504	0,0754	0,0000	0,0850	0,051	0,033	0,011
psbF	0,0000	0,0000	0,0825	0,0000	0,0000	0,0000	0,0634	0,3859	0,3742	0,0000	0,091	0,156	0,049
psbH	0,1757	0,1333	0,2176	0,1338	0,1310	0,1260	0,0909	0,1957	0,1279	0,0681	0,140	0,045	0,014
psbI	0,0000	0,0687	0,0000	0,0000	0,0495	0,0556	0,0546	0,0665	0,0000	0,0000	0,030	0,032	0,010
psbJ	0,0940	0,6152	0,0906	0,2970	0,4515	0,0969	0,1360	0,1566	0,0886	-	0,225	0,191	0,060
psbK	0,2796	0,1515	0,2890	0,2205	0,2784	0,1605	0,1472	0,1676	0,3063	0,4355	0,244	0,092	0,029
psbL	0,0882	0,0396	0,0717	0,0000	0,1208	0,0000	0,0000	0,0000	0,0501	0,0000	0,037	0,045	0,014
psbM	0,0971	0,0545	0,0971	0,0257	0,0370	0,0431	0,0545	0,0774	0,0314	0,0572	0,057	0,026	0,008
psbN	0,0755	0,2839	0,0574	0,1518	0,2839	0,2676	0,2125	0,1422	0,0000	0,0000	0,147	0,112	0,035
psbT	0,0329	0,2752	0,0275	0,0988	0,3309	0,2079	0,4308	0,9361	0,0891	0,3568	0,279	0,271	0,086
psbZ	0,1769	0,0318	0,1669	0,1400	0,0547	0,0285	0,0540	0,0766	0,0488	0,0367	0,081	0,057	0,018
rbcL	0,0411	0,0589	0,0551	0,0566	0,0558	0,0223	0,0502	0,0483	0,0309	0,0520	0,047	0,012	0,004
rpl14	0,0819	0,0998	0,0901	0,0906	0,1043	0,0923	0,1432	0,1121	0,0978	0,1099	0,102	0,017	0,005
rpl16	0,0369	0,1747	0,0316	0,0336	0,1770	0,1829	0,1763	0,1619	0,0514	0,0520	0,108	0,071	0,022
rpl20	0,2044	0,5737	0,2707	0,2056	0,5499	0,3537	0,4977	0,4063	0,2421	0,1702	0,347	0,152	0,048
rpl23	0,9922	-	0,8093	0,4272	-	0,3193	0,3880	0,4096	1,6667	0,5322	0,693	0,457	0,144
rpl2	0,1640	0,2065	0,1626	0,1291	0,1701	0,2848	0,2490	0,1089	0,2293	0,1516	0,186	0,055	0,018
rpl32	0,2465	0,1849	0,3421	0,5965	0,2371	0,2983	-	0,3965	0,1849	0,3079	0,311	0,128	0,041
rpl33	-	0,1836	-	0,3174	0,1713	0,2975	0,4420	0,3761	0,3619	0,1258	0,284	0,113	0,036
rpl36	0,1335	0,5344	0,0955	0,0170	0,3199	0,1811	0,1115	0,1075	0,0583	0,0613	0,162	0,155	0,049
rpoA	0,3190	0,3132	0,2900	0,2564	0,4758	0,2233	0,2952	0,2490	0,2916	0,2102	0,292	0,074	0,023
rpoB	0,1511	0,1462	0,1527	0,1429	0,2089	0,1410	0,1736	0,1414	0,1420	0,1774	0,158	0,022	0,007
rpoC1	0,1567	0,1844	0,1506	0,1318	0,2427	0,1595	0,1781	0,1370	0,1282	0,2095	0,168	0,037	0,012
rpoC2	0,2353	0,2584	0,2293	0,2081	0,2625	0,3211	0,2800	0,2590	0,2116	0,2320	0,250	0,034	0,011
rps11	0,1044	0,1989	0,0999	0,0922	0,3413	0,1787	0,2975	0,2336	0,0863	0,0987	0,173	0,093	0,029
rps12	0,0208	0,1047	0,0000	0,0000	0,1469	0,0750	0,0479	0,0248	0,0420	0,0000	0,046	0,049	0,016
rps14	0,1077	0,3402	0,1310	0,1141	0,6504	0,3264	0,4024	0,3649	0,1381	0,2968	0,287	0,171	0,054
rps15	0,3496	0,3447	0,4288	0,1856	0,4003	0,5545	0,2653	0,2684	0,2270	0,4483	0,347	0,114	0,036
rps16	-	-	-	0,3348	-	-	-	-	0,1955	0,2711	0,267	0,070	0,022
rps18	0,4023	0,3232	0,3942	0,3120	0,3671	0,5974	-	0,8821	0,1750	0,4451	0,433	0,203	0,064
rps19	0,2344	0,3145	0,2344	0,1951	0,1081	0,2017	0,0829	0,1627	0,1984	0,2717	0,200	0,070	0,022
rps2	0,2001	0,1230	0,1719	0,1240	0,1796	0,1378	0,2240	0,0911	0,1449	0,1237	0,152	0,041	0,013
rps3	0,1659	0,7449	0,1874	0,2218	0,8166	0,2594	0,4048	0,3771	0,1155	0,1199	0,341	0,252	0,080
rps4	0,2123	0,2754	0,2094	0,2224	0,3207	0,2576	0,2527	0,2700	0,2391	0,2954	0,256	0,036	0,011
rps7	0,1959	0,3582	0,1584	0,0791	0,8269	0,2118	0,5665	0,6690	0,0684	0,0669	0,320	0,275	0,087
rps8	0,2010	0,1968	0,2125	0,1355	0,2658	0,2191	0,2460	0,2104	0,3012	0,2356	0,222	0,044	0,014
ycf1	0,7313	0,5294	0,6155	0,5643	0,5616	0,7320	-	0,6294	0,5280	0,5166	0,601	0,083	0,026
ycf2	0,7391	0,6354	0,7400	0,6402	0,6920	0,6680	0,6361	0,6427	0,6809	0,7410	0,682	0,045	0,014
ycf3	0,1042	0,0550	0,0982	0,1136	0,0513	0,0405	0,0644	0,0595	0,0456	0,1010	0,073	0,028	0,009
ycf4	0,4633	-	0,4146	-	0,6444	-	-	-	0,8406	0,8802	0,649	0,212	0,067

Table S8: Identification of the presence of inverted repeated elements in *trnS-GGA* and *trnS-GCU* genes within rosid plastomes.

					Rosid Species	**Presence (length) of inverted repeat elements in trnS-GGA and trnS-GCU**	**Position of trnS-GGA**	**Position of trnS-GCU**	**Distance between trnS-GGA and trnS-GCU**
Eurosids I (Fabids)	Cucurbitales	Cucurbitaceae			*Cucumis sativus*	11	47328-47412	8692-8779	38549
	Fabales	Fabaceae	Papilionideae (non IRLC)	Genistoideae	*Lupinus luteus*	29	52025-52112	16025-15939	36000
				Phaseoleae	*Glycine max*	29	16174-16087	53214-53300	37040
					Phaseolus vulgaris	23	18936-18847	55334-55420	34398
					Vigna radiata	23	18917-18830	55910-55996	36993
				Loteae	*Lotus japonicus*	29	16150-16063	52945-53031	36795
			Papilionideae (IRLC)	Cicereae	*Cicer arietinum*	29	15990-15903	53266-53352	37276
				Trifolieae	*Medicago truncatula*	29	108094-108181	71071-70985	37023
					Trifolium subterraneum	29	117904-117991	40150-40065	77754
				Fabeae	*Pisum sativum*	no	15872-15785	74373-74287	58415
					Lathyrus sativus	29	15578-15491	66790-66876	51212
	Malpighiales	Euphorbiaceae			*Manihot esculenta*	23	47628-47714	8899-8812	38729
		Salicaceae			*Populus alba*	29	44930-45016	6898-6811	38032
	Rosales	Moraceae			*Morus indica*	29	47339-47425	8933-8846	38406
Eurosids II (Malvids)	Brassicales	Brassicaceae			*Arabidopsis thaliana*	28	44827-44913	7785-7872	36955
	Sapindales	Rutaceae			*Citrus sinensis*	23	48270-48356	8818-8731	39452
	Geraniales	Geraniaceae			*Pelargonium x. hortorum*	29	49703-49789	7503-7416	42200
	Myrtales	Myrtaceae			*Eucalyptus globulus*	29	48966-49052	8996-8909	39970
		Onagraceae			*Oenothera biennis*	29	19709-19623	58225-58307	38516
	Vitales	Vitaceae			*Vitis vinifera*	29	48831-48917	9528-9441	39303

The first complete chloroplast genome of the genistoid legume *Lupinus luteus*: Evidence for a novel major lineage-specific rearrangement and new insights on plastome evolution in the Legume family.

Authors

Guillaume E. Martin[1,2†], Mathieu Rousseau-Gueutin[1†], Solenn Cordonnier[1], Oscar Lima[1], Sophie Michon-Coudouel[3], Delphine Naquin[1,4], Julie Ferreira de Carvalho[1], Malika Aïnouche[1], Armel Salmon[1], Abdelkader Aïnouche[1*]

[1]. UMR CNRS 6553 Ecobio, OSUR (Observatoire des Sciences de l'Univers de Rennes), Université de Rennes 1 / Université Européenne de Bretagne, 35 042 Rennes, France

[2]. Current address: CIRAD (Centre de coopération Internationale en Recherche Agronomique pour le Développement), UMR AGAP, F-34398 Montpellier, France

[3]. Plate-forme Génomique Environnementale et Fonctionnelle, OSUR-CNRS, Université de Rennes 1, 35042 Rennes, France

[4]. Plate-forme de Bioinformatique, Genouest INRIA/IRISA, Université de Rennes-1, 35042 Rennes, France

[5]. Current address: Plate-forme IMAGIF, FRC3115 CNRS, 91198 Gif sur Yvette Cedex, France

[*]. Corresponding author: Dr. Abdelkader Aïnouche

UMR CNRS 6553 University of Rennes 1, 35 042 Rennes, France

Phone: +33 (0)2 23 23 61 19

Fax: +33 (0)2 23 23 50 47

email address: kader.ainouche@univ-rennes1.fr

[†]. The two first authors contributed equally.

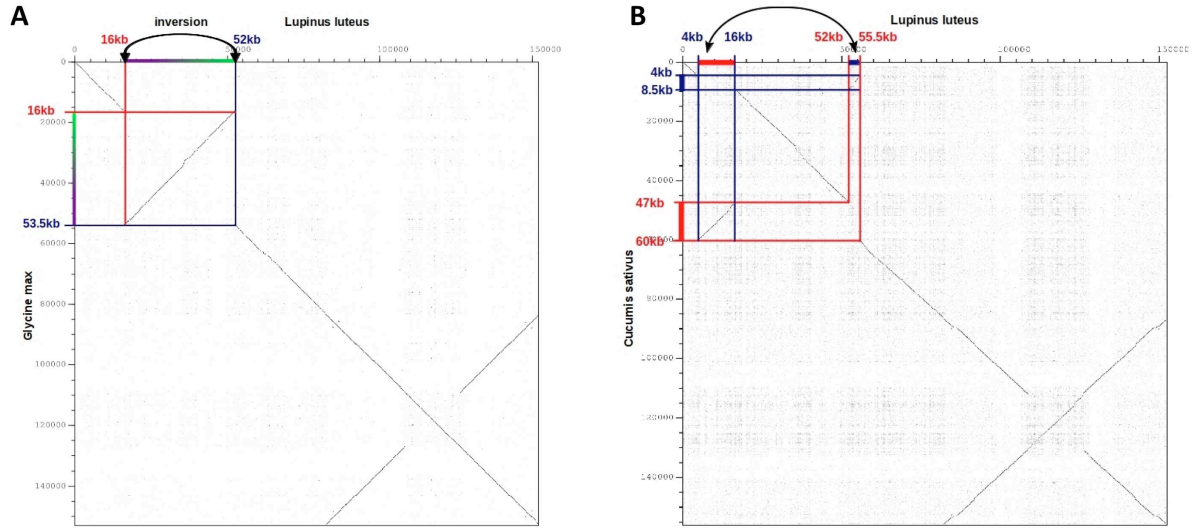

Fig. S1. Dot matrix plots showing the presence of a 36 kb inversion in *Lupinus luteus* plastome. (A) Comparison of *Lupinus luteus* to *Glycine max* (Fabaceae) plastome revealing the presence of a 36 kb inversion. (B) Comparison of *Lupinus luteus* to *Cucumis sativus* (Fabaceae outgroup) revealing the presence of a 36 kb inversion embedded in the 50 kb inversion present in most Papilionideae (Fabaceae).

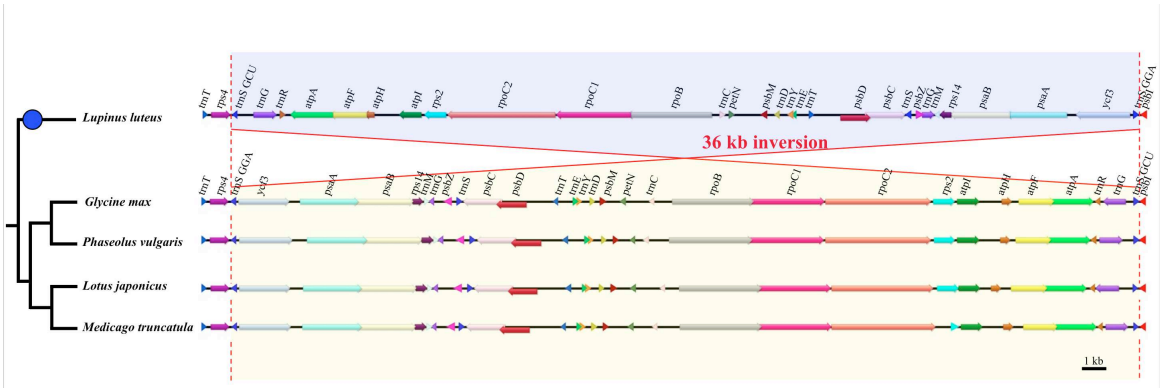

Fig. S2. Comparative plastomic maps showing the presence of a 36 kb inversion in *Lupinus luteus* in comparison to other Papilionoideae. The genes present and directly surrounding this inversion are shown. Only four representatives of the Fabaceae species having their plastomes fully sequenced (*Glycine max*, *Phaseolus vulgaris*, *Lotus japonicus* and *Medicago truncatula*) are presented here and the phylogenetic relationships between the different Papilionoideae represented are redrawn from Cardoso et al. (2012). The blue circle on a branch indicates the origin of the 36kb inversion event. The partial plastomic maps are drawn to scale.

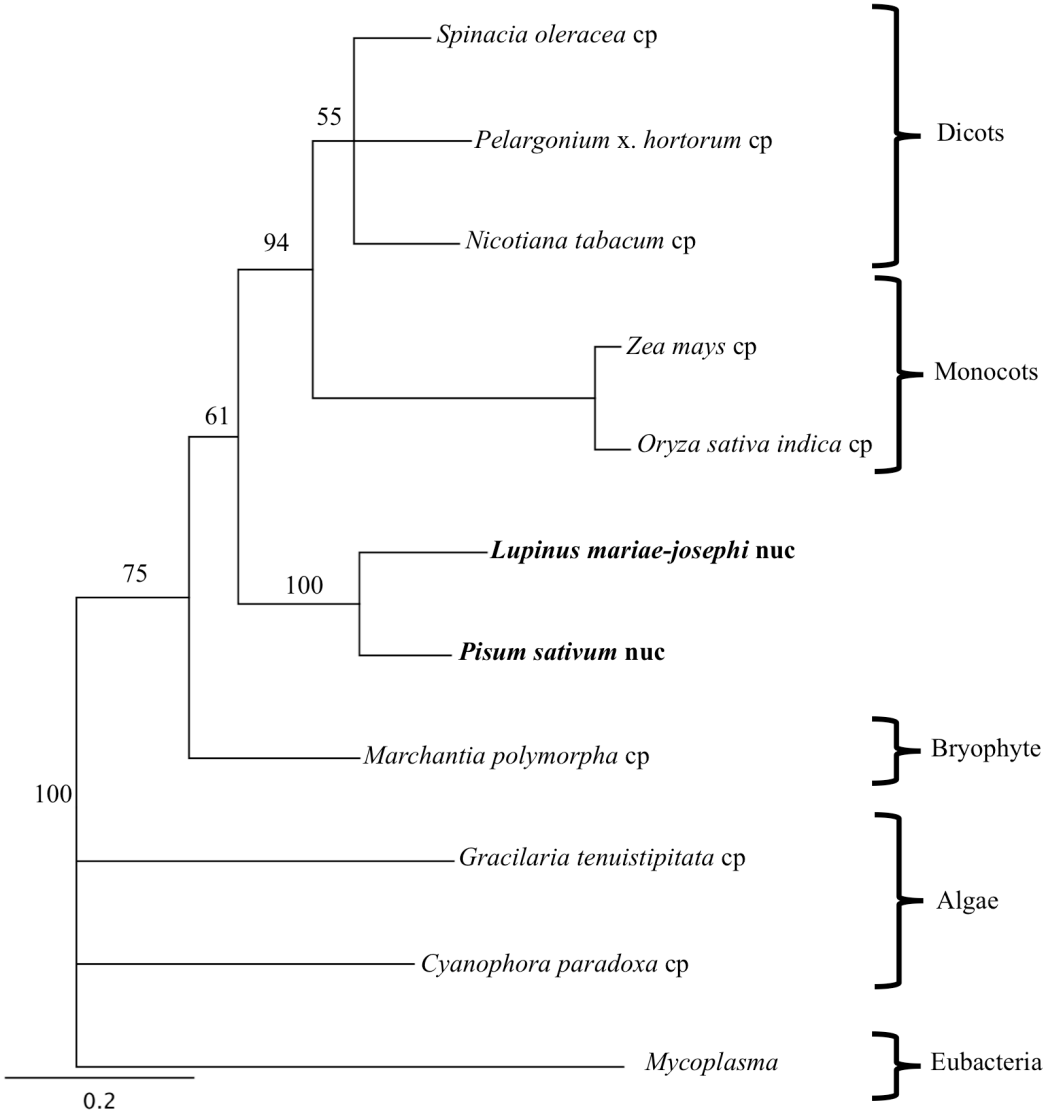

Fig. S3. Phylogenetic analysis of plastidic and nuclear *rpl22* protein sequences. Presented is a Neighbor Joining tree obtained using Jukes-Cantor model and rooted using the eubacteria *Mycoplasma*. Bootstrap values obtained from 10,000 replicates are shown above the branches. The scale bar denotes substitutions per site along the branches. GenBank accession numbers of the *rpl22* chloroplastic sequences are as follows: *Cyanophora paradoxa* (NC_001675), *Gracilaria tenuistipitata* (NC_006137), *Marchantia polymorpha* (NC_001319) *Mycoplasma* (M74770), *Nicotiana tabacum* (NC_001879), *Oryza sativa indica* (NC_008155), *Pelargonium* x. *hortorum* (NC_008454), *Spinacia oleracea* (NC_002202) and *Zea mays* (NC_001666). GenBank accession numbers of the *rpl22* genes functionally transferred from the chloroplast to the nucleus in a common ancestor of all flowering plants (Gantt *et al.*, 1991) are: *Lupinus mariae josephi* nuc (sequence available upon request) and *Pisum sativum* nuc (M60951).

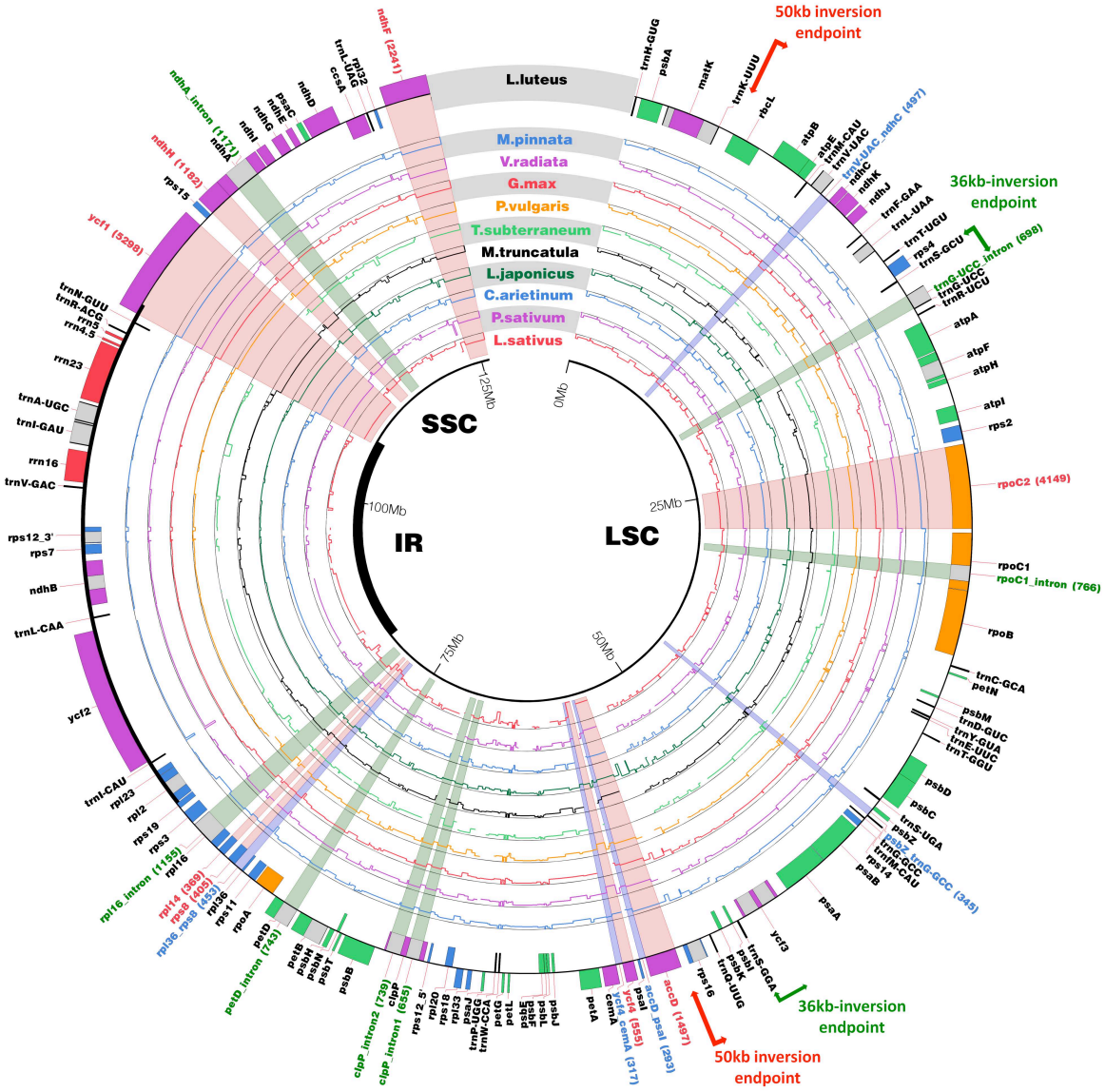

Fig. S4. Pairwise distance between *Lupinus luteus* and other Fabaceae orthologous plastomic regions. Pairwise distance was calculated using K2p model (Kimura, 1980) for introns and intergenic spacers, whereas sequence divergence of protein-coding genes (exons) were calculated using the synonymous (Ks) nucleotide substitution rates. The protein-coding, intronic or intergenic regions presenting a higher evolutionary rate than those previously used in Fabaceae evolutionary studies (such as *rbcL* gene, *trnK-UUU* intron and *trnL-trnT* IGS) and a minimum size of 300 bp are highlighted in red (exons), green (introns) and blue (IGS). The size (in bp) of each highlighted region in *Lupinus* is indicated between brackets. Only one IR region is represented on each plastomic map. The endpoints of the 50 kb and 36 kb inversions are indicated by red and green arrows, respectively.

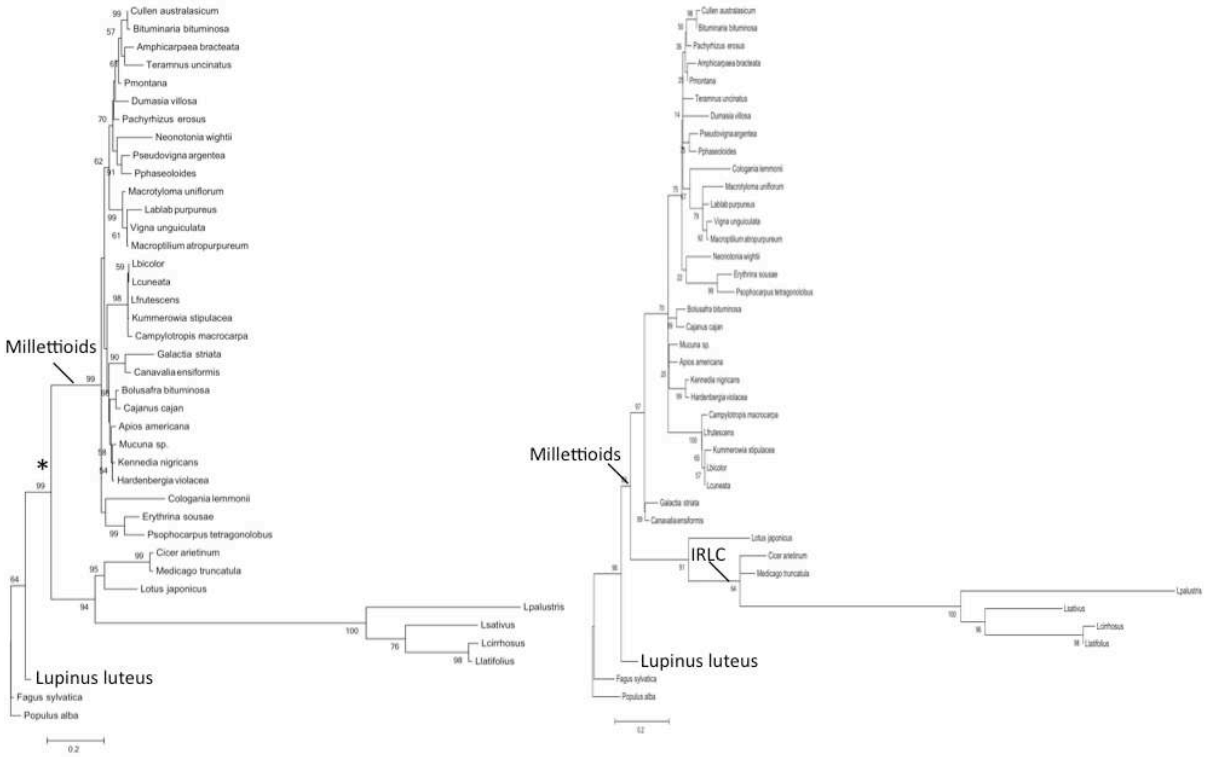

Fig. S5. Synonymous and nonsynonymous divergence in Legume *ycf4* gene. Shown are *dN* (left) and *dS* (right) phylogenetic trees obtained from a codon based maximum likelihood method (GTR + G + I model). The trees, obtained using MEGA 5.0 (Tamura, 2011), were rooted with the *Populus alba* sequence. Bootstrap values obtained from 1,000 replicates are shown on each branch. The asterisk indicates the branch in which rate acceleration is first seen (Magee *et al.*, 2010). The scale bar denotes substitutions per site along the branches. The legume *ycf4* sequences were obtained from Stefanovic *et al.* (2009) and Magee *et al.* (2010).